

**Kaupallisen tekstin- ja tiedonlouhinnan asema EU:n muuttuneessa
tekijänoikeussäätelyssä**

Helsingin yliopisto
Oikeustieteellinen tiedekunta
Vaasan oikeustieteellinen koulutus
Kauppaoikeus
OTM-tutkielma
Huhtikuu 2021
Tekijä: Juho Aittasalo
Ohjaaja: Marcus Norrgård

Tiivistelmä

Tiedekunta: Oikeustieteellinen tiedekunta

Koulutusohjelma: Oikeustieteen maisterin koulutusohjelma

Opintosuunta: Kauppa-oikeus

Tekijä: Juhon Aittasalo

Työn nimi: Kaupallisen tekstin- ja tiedonlouhinnan asema EU:n muuttuneessa tekijänoikeussäätelyssä

Työn laji: OTM-tutkielma

Kuukausi ja vuosi: Huhtikuu 2021

Sivumäärä: XV + 71

Avainsanat: Immateriaalioikeus, tekijänoikeus, DSM-direktiivi, tekstin- ja tiedonlouhinta

Säilytyspaikka: Helsingin yliopiston kirjasto

Muita tietoja:

Tiivistelmä:

Tämän tutkielman kohde on kaupallisen tekstin- ja tiedonlouhinnan suhde tekijänoikeussäätelyyn. Suhdetta tarkastellaan pääasiallisesti Euroopan unionin (EU) tekijänoikeudesta ja lähioikeuksista digitaalisilla sisämarkkinoilla hyväksymän direktiivin (DSM-direktiivi) 4 artiklan kautta. Artiklaa peilataan sekä EU:n aiempaan oikeustilaan että Yhdysvaltojen relevanttiin tekijänoikeussäätelyyn ja -traditioon. Tutkielman kohde on ajankohtainen, koska tekstin- ja tiedonlouhintaan perustuvat teknologiat mahdollistavat erilaisten kaavojen ja säännönmukaisuuksien havaitsemisen valtavista datamassoista.

EU:n tekstin- ja tiedonlouhintaa koskeva säätelykehikko oli pitkään rakentunut tietoyhteiskuntadirektiivin, tietokantadirektiivin ja tietokoneohjelmadirektiivin varaan. Missään näistä direktiiveistä ei kuitenkaan käsitelty nimenomaisesti tekstin- ja tiedonlouhintaa. Tekstin- ja tiedonlouhinnan tekijänoikeudellista sallittavuutta tarkasteltiinkin tällöin lähinnä tietoyhteiskuntadirektiivin kappaleen valmistamista koskevan oikeuden ja siihen asetetun pakollisen poikkeuksen tai rajoituksen kautta. Poikkeuksen tai rajoituksen soveltumisen lukuisat ja luonteeltaan kumulatiiviset edellytykset asettivat ei-kaupallisen tekstin- ja tiedonlouhinnan tekijänoikeudellisen sallittavuuden epävarmaksi. Kaupallisen louhinnan ei puolestaan voida katsoa mahtuneen poikkeuksen tai rajoituksen raamien sisälle lainkaan.

DSM-direktiivin 4 artiklan 1 kohta velvoittaa jäsenvaltiot säätämään tekijänoikeuden poikkeuksesta tai rajoituksesta, kun kyseessä on kappaleen valmistaminen tai kopiointi tekstin- ja tiedonlouhintaa varten teoksista ja muusta suojatusta aineistosta, joihin on laillinen pääsy. Poikkeus tai rajoitus kohdistuu kaikkiin edellisessä kappaleessa mainittuihin direktiiveihin ja koskee DSM-direktiivin 3 artiklasta poikkeavasti muita tekstin- ja tiedonlouhinnan tarkoituksia kuin tieteellistä tutkimusta. Näin ollen siitä voidaan puhua myös kaupallista tekstin- ja tiedonlouhintaa koskevana säännöksenä. 4 artiklan 3 kohta kuitenkin rajoittaa artiklan sovellettavuutta antamalla oikeudenhaltijoille mahdollisuuden pidättää oikeutensa 1 kohdassa tarkoitettuun teosten ja muun suojatun aineiston käyttöön. Oikeudenhaltijoiden käyttäessä mahdollisuuttaan aktiivisesti artiklan merkitys ja muutosvoima aiempaan oikeustilaan nähden voi hälvetä.

Yhdysvalloissa tekijänoikeuden rajoitukset ja poikkeukset perustuvat yksittäisten säännösten sijaan fair use -doktriiniin. Tällä taataan joustavampi ja teknologianeutraalimpi lähestymistapa rajoituksiin ja poikkeuksiin. Kaupallinen tekstin- ja tiedonlouhinta onkin Google Books -tapauksen yhteydessä todettu kohtuulliseksi käytöksi muun muassa käytön transformatiivisen tarkoituksen vuoksi. Tuomioistuinten tapauksessa tekemä fair use -kriteereiden punninta osoittaa, että myös EU:ssa olisi voinut olla tilausta yleistä ja yksityistä etua paremmin tasapainottavalle kaupallista tekstin- ja tiedonlouhintaa koskevalle poikkeukselle tai rajoitukselle.

SISÄLLYS

LÄHTEET	III
LYHENTEET.....	XV
1. JOHDANTO.....	1
1.1 JOHDATUS AIHEESEEN.....	1
1.2 TUTKIMUSKOhteET JA -KYSYMYKSET.....	5
1.3 TUTKIMUSMENETELMÄ JA TUTKIMUKSEN RAKENNE	7
2. TEKSTIN- JA TIEDONLOUHINNASTA	11
2.1 MITÄ ON TEKSTIN- JA TIEDONLOUHINTA?.....	11
2.2 TEKSTIN- JA TIEDONLOUHINTAPROSESSIN PÄÄPIIRTEET	14
2.3 TEKSTIN- JA TIEDONLOUHINNAN SUHDE KONEOPPIMISEEN JA TEKÖÄLYYN.....	18
3. TEKSTIN- JA TIEDONLOUHINNAN SÄÄNTELY EU:SSA	22
3.1 DSM-DIREKTIIVIÄ EDELTÄNYT OIKEUSTILA	22
3.1.1 Tietoyhteiskuntadirektiivi	22
3.1.2 Tietokantadirektiivi	31
3.1.3 Tietokoneohjelmadirektiivi.....	33
3.2 DSM-DIREKTIIVIÄ SEURAAVA OIKEUSTILA	34
3.2.1 DSM-strategiasta tekijänoikeuden ja tekstin- ja tiedonlouhinnan kannalta.....	34
3.2.2 DSM-direktiivin 4 artiklan sisältö ja perustelut.....	36
3.2.3 Tekstin- ja tiedonlouhinta kaupallisessa tarkoituksessa: Mikä muuttuu?.....	39
4. TEKSTIN- JA TIEDONLOUHINNAN SÄÄNTELY YHDYSVALLOISSA	48
4.1 RELEVANTIN TEKIJÄNOIKEUSSÄÄNTELYN LÄHTÖKOHDISTA	48
4.2 <i>GOOGLE BOOKS</i> -TAPAUS: TEKSTIN- JA TIEDONLOUHINTA ON KOHTUULLISTA KÄYTTÖÄ	50
4.2.1 <i>Google Books Library Project</i>	50
4.2.2 <i>Google Books</i> -tapauksen käsittely.....	51
4.3 FAIR USE -KRITERIEN ARVIOINTIA KAUPALLISEN TEKSTIN- JA TIEDONLOUHINNAN NÄKÖKULMASTA ..	59
4.4 VERTAILUA EU:N OIKEUSTILAAN NÄHDEN.....	62
5. JOHTOPÄÄTÖKSIÄ.....	65

Lähteet

Kirjallisuus ja muut julkaisut

Aggarwal – ChengXiang 2012

Aggarwal, Charu C.; Zhai, ChengXiang: An Introduction to Text Mining. Teoksessa Aggarwal, Charu C.; Zhai, ChengXiang (toim.): Mining Text Data. Springer Science + Business Media LLC 2012.

Aggarwal 2015

Aggarwal, Charu C.: Data Mining: The Textbook. Springer International Publishing Switzerland 2015.

Alpaydin 2016

Alpaydin, Ethem: Machine Learning: The New AI. MIT Press, Cambridge, Massachusetts 2016.

Alpaydin 2014

Alpaydin, Ethem: Introduction to Machine Learning. The MIT Press, Cambridge, Massachusetts 2014.

Bentham 1891

Bentham, Jeremy: Fragment on Government. Oxford, Clarendon Press 1891.

Berkhin 2006

Berkhin P.: A Survey of Clustering Data Mining Techniques. Teoksessa Grouping Multidimensional Data: Recent Advances in Clustering (Jacob Kogan; Charles Nicholas; Marc Teboulle), Springer-Verlag Berlin Heidelberg 2006.

Bottis ym. 2019

Bottis, Maria; Papadopoulos, Marinos; Zampakolas, Christos; Ganatsiou, Paraskevi: Text and Data Mining in the EU “Acquis Communautaire” Tinkering with TDM & Digital Legal Deposit. Teoksessa Erasmus Law Review, vol. 12, no. 2, 2019, s. 190-208.

Bramer 2016

Bramer, Max: Principles of Data Mining. Springer-Verlag London Ltd. 2016.

Caspers – Guibault 2016

Caspers, Marco; Guibault, Lucie: Deliverable 3.3 Baseline report of policies and barriers of TDM in Europe. FutureTDM 2016.

Diaz 2013

Diaz, Angel Siegfried: Fair Use & Mass Digitization: The Future of Copy-Dependent Technologies after Authors Guild v. HathiTrust. Teoksessa Berkeley Technology Law Journal, vol. 28, no. Annual Review Issue, 2013, s. 683-714.

Ducato – Strowel 2019

Ducato, Rossana; Strowel, Alain: Limitations to Text and Data Mining and Consumer Empowerment. Making the Case for a Right to “Machine Legibility”. Kluwer Copyright Blog 2019. <URL: http://copyrightblog.kluweriplaw.com/2019/03/19/limitations-to-text-and-data-mining-and-consumer-empowerment-making-the-case-for-a-right-to-machine-legibility/?doing_wp_cron=1593079100.1405661106109619140625>. Viitattu 25.6.2020.

Dusollier 2020

Dusollier, Séverine: The 2019 Directive on Copyright in the Digital Single Market: Some progress, a few bad choices, and an overall failed ambition. Teoksessa Common Market Law Review, vol. 57, issue 4, 2020, s. 979-1030.

Ertel 2017

Ertel, Wolfgang: Introduction to Artificial Intelligence. Springer International Publishing AG 2017.

Ervo 2008

Ervo, Laura: Virallisen vastaväittäjän lausunto Antti Tapanilan väitöskirjasta. Lakimies 2008 s. 316-327.

Flynn – Geiger – Quintais ym. 2020

Flynn, Sean; Geiger, Christopher; Quintais, João Pedro; Margoni, Thomas; Sag, Matthew; Guibault, Lucie; Carroll, Michael: Implementing User Rights for Research in the Field of Artificial Intelligence: a Call for International

Action. Teoksessa European Intellectual Property Review 2020, 42(7), s. 393-398. Sweet & Maxwell and its Contributors.

Geetha ym. 2019

Geetha N; Kishan Das Menon; Achal Raj Jain; Janardhan V; Dr. Piyush Kumar Pareek: Quantitative Analysis of Student Data Mining. Teoksessa Proceedings of Second International Conference on Emerging Trends in Science & Technologies for Engineering Systems (ICETSE-2019) SJC Institute of Technology, Chickballapur, Karnataka, India 17th and 18th May 2019.

Geiger 2006

Geiger, Christophe: Copyright and Free Access to Information: For a Fair Balance of Interests in a Globalised World. Teoksessa European Intellectual Property Review 2006, 28(7), s. 366-373.

Geiger 2010

Geiger, Christophe: The Future of Copyright in Europe: Striking a Fair Balance Between Protection and Access to Information. Teoksessa Intellectual Property Quarterly 2010, 1, s. 1-14.

Geiger – Frosio – Bulayenko 2018

Geiger, Christophe; Frosio, Giancarlo; Bulayenko, Oleksandr: The Exception for Text and Data Mining in the Proposed Directive on Copyright in the Digital Single Market – Legal Aspects. Teoksessa: Center for International Intellectual Property Studies Research Paper No. 2018-02, Université de Strasbourg.

Geiger – Izyumenko 2019

Geiger, Christophe; Izyumenko Elena: Towards a European Fair Use Grounded in Freedom of Expression. Teoksessa: American University International Law Review, 35(1), 1-74.

Geiger – Schönherr 2014

Geiger, Christophe; Schönherr, Franciska: The Information Society Directive. Teoksessa Stamatoudi, A. Irini; Torremans, Paul (toim.): EU Copyright Law, Edwar Elgar Publishing, 2014, s. 395-527.

Gervais 2017

Gervais, Daniel J.: (Re)structuring Copyright: A Comprehensive Path to International Copyright Reform. Edward Elgar Publishing 2017.

Ginsburg 2020

Ginsburg, Jane C.: Fair Use in the United States: Transformed, Deformed, Reformed? Teoksessa: Singapore Journal of Legal Studies, 2020 Sing. J. Legal Stud. 265, s. 265-295.

Gollapudi 2016

Gollapudi, Sunila: Practical Machine Learning: Tackle the real-world complexities of modern machine learning with innovative and cutting-edge techniques. Packt Publishing 2016.

Guze 2019

Guze, Nathan: Net Neutrality and the European Union's Copyright Directive for the Digital Single Market. Loyola of Los Angeles International and Comparative Law Review, vol. 43, no. 1, 2019-2020, s. 63-vi.

Haarmann 2014

Haarmann, Pirkko-Liisa: Immateriaalioikeus, Talentum 2014, Helsinki.

Haarmann 2005

Haarmann, Pirkko-Liisa: Tekijänoikeus ja lähioikeudet, Talentum 2005, Helsinki.

Han – Kamber – Pei 2011

Han, Jiawei; Kamber, Micheline; Pei, Jian: Data Mining: Concepts and Techniques. Elsevier Science & Technology 2011.

Hirvonen 2011

Hirvonen, Ari: Mitkä metodit? Opas oikeustieteen metodologiaan. Yleisen oikeustieteen julkaisuja 17, Helsinki 2011.

Hua 2014

Hua, Jerry Jie: Toward a More Balanced Approach: Rethinking and Readjusting Copyright Systems in the Digital Network Era. Springer-Verlag Berlin Heidelberg, 2014.

Hugenholtz 2019

Hugenholtz, Bernt: The New Copyright Directive: Text and Data Mining (Articles 3 and 4). Kluwer Copyright Blog. <URL: http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/?doing_wp_cron=1591013316.1060500144958496093750> Viitattu: 25.6.2020.

Hughes 2017

Hughes, Justin: Fair Use and Its Politics – at Home and Abroad. Teoksessa Okediji, Ruth L. (toim.): Copyright Law in an Age of Limitations and Exceptions. Cambridge University Press 2017, s. 234-274.

Husa 2010

Husa, Jaakko: Valkoista yksisarvista pyydystämässä vai mörköä paossa – ”oikeaa oikeusvertailua”? Lakimies 5/2010.

Husa ym. 2008

Husa, Jaakko; Mutanen, Anu; Pohjalainen, Teuvo: Kirjoitetaan juridiikkaa – Ohjeita oikeustieteellisten kirjallisten töiden laatijoille. Talentum, Helsinki 2008.

Jongsma 2019

Jongsma, Daniël: Creating EU Copyright Law: Striking a Fair Balance. Publications of the Hanken School of Economics, Nr 334, Helsinki 2019.

Jordan ym. 2015

Jordan, M.I.; Mitchell, T.M.: Machine Learning: Trends, perspectives, and prospects. Julkaisussa Science 349(6245) 2015.

Kaczynski ym. 2016

Kaczynski, David; Gandy, Lisa; Hu, Gonghzu: Innovations in News Media: Crisis Classification System. Teoksessa Advances in Data Mining: Applications and Theoretical Aspects: 16th Industrial Conference, ICDM 2016 New York, NY, USA, July 13-17, 2016 Proceedings (Petra Pernert), Springer International Publishing Switzerland 2016.

Kaplow 1992

Kaplow, Louis: Rules Versus Standards: An Economic Analysis. Teoksessa Duke Law Journal, vol. 42, no. 3, 1992, s. 557-629.

Kolehmainen 2015

Kolehmainen, Antti: Tutkimusongelma ja metodi lainopillisessa työssä, teoksessa Oikeustieteellinen opinnäyte – Artikkeleita oikeustieteellisten opinnäytteiden vaatimuksista, metodista ja arvostelusta (toim. Tarmo Miettinen). Edilex 2015/29.

Koster 1994

Koster, Martijn: A Standard for Robot Exclusion. <URL: <http://www.robotstxt.org/orig.html>>. Viitattu 30.3.2021.

Larose, D. – Larose, C. 2014

Larose, Daniel T., Larose Chantal D.: Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons Inc., 2014.

Leval 1990

Leval, Pierre N.: Toward a Fair Use Standard. Teoksessa Harvard Law Review, vol. 103, no. 5, 1990, s. 1105-1136.

Matulionyte 2015

Matulionyte, Rita: 10 Years for Google Books and Europeana: Copyright Law Lessons That the EU Could Learn from the USA. Teoksessa International Journal of Law and Information Technology, vol. 24, no. 1, 2015, s. 44-71.

Mazziotti 2008

Mazziotti, Giuseppe: EU Digital Copyright Law and the End-User. Springer-Verlag Berlin Heidelberg, 2008.

Murphy 2012

Murphy, Kevin P.: Machine Learning: A Probabilistic Perspective. MIT Press, Cambridge 2012.

Määttä 2015

Määttä, Tapio: Metodinen pluralismi oikeustieteessä – ympäristöoikeudellisen tutkimuksen suuntaukset ja menetelmät. Edilex 2015/45.

Oesch 2017

Oesch, Rainer: Johdatus aiheeseen: Yleinen etu ja immateriaalioikeuden suoja – Mitä uutta? Teoksessa Oesch, Rainer; Eloranta, Mikko; Heino, Mari; Kokko, Mira (toim.): Immateriaalioikeudet ja yleinen etu, Alma Talent Oy, Helsinki 2017, s. 1-20.

Pearl 2019

Pearl, Judea: The Seven Tools of Causal Inference with Reflections on Machine Learning. Teoksessa Communications of the ACM, 62(3), 2019, s. 54-60.

Panezi 2018

Panezi, Argyri: Europe's New Renaissance: New Policies and Rules for Digital Preservation and Access to European Cultural Heritage. Teoksessa Columbia Journal of European Law, vol. 24, no. 3, 2018, s. 596-611.

Pihlajarinne 2004

Pihlajarinne, Taina: Immateriaalioikeudet. Teoksessa Villa, Seppo; Airaksinen, Manne; Bärlund, Johan; Jauhiainen, Jyrki; Kaisanlahti, Timo; Knuts, Mårten; Kuoppamäki, Petri; Kymäläinen, Seppo; Mähönen, Jukka; Pihlajarinne, Taina; Raitio, Juha; Rissanen, Kirsti; Viitanen, Klaus; Wilhelmsson, Thomas: Yritysoikeus, Alma Talent Oy, Helsinki 2014, luku V.

Quintais 2020

Quintais, João Pedro: The new copyright in the Digital Single Market Directive: a critical look. Teoksessa European Intellectual Property Review 2020, 42(1), s. 28-41. Sweet & Maxwell and its Contributors.

Reinsel ym. 2018

Reinsel, David; Gantz, John; Rydning, John: Data Age 2025, The Digitization of the World – From Edge to Core, An IDC White Paper 2018.

Renda 2017

Renda, Andrea: Will the DSM Strategy Spur Innovation? Teoksessa Inter economics, 2017-07, vol. 52 (4), s. 197-201. Springer Berlin Heidelberg.

Rosati 2019

Rosati, Eleonora: Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity.

Sag 2012

Sag, Matthew: Predicting Fair Use. *Teoksessa Ohio State Law Journal*, vol. 73, no. 1, 2012, s. 47-92.

Sag 2019

Sag, Matthew: The New Legal Landscape for Text Mining and Machine Learning. *Teoksessa Journal of the Copyright Society of the USA*, vol. 66, no. 2, s. 291-(vi).

Schneider 1998

Schneider, Mark: The European Union Database Directive. *Teoksessa Berkeley Technology Law Journal*, vol. 13, no. 1, 1998, s. 551-654.

Shapiro 2019

Shapiro, Ted: Directive 2001/29/EC on Copyright in the Information Society. *Teoksessa Lindner, Brigitte; Shapiro, Ted (toim.): Copyright in the Information Society, A Guide to National Implementation of the European Directive, Second Edition. Edward Elgar Publishing 2019*, s. 40-125.

Thau 1995

Thau, Stephen B.: Copyright, Privacy and Fair Use. *Hofstra Law Review*, vol. 24, no. 1, 1995, s. 179-222.

Triaille 2008

Triaille, Jean-Paul: La question des copies caches et la responsabilité des intermédiaires. *Teoksessa: Strowel, Alain – Triaille, Jean-Paul (toim.): Google et les nouveaux services en ligne. Larcier, 2008*.

Triaille ym. 2014

Triaille, Jean-Paul - d'Argenteuil, Jérôme de Meeûs – de Francquen, Amélie: Study on the legal framework of text and data mining (TDM), European Commission, 2014.

Warwick 2011

Warwick, Kevin: Artificial Intelligence: The Basics. Taylor & Francis Group, 2011.

Xalabarder 2014

Xalabarder, Raquel: Google Books and Fair Use: A Tale of Two Copyrights. Teoksessa Journal of Intellectual Property, Information Technology and Electronic Commerce Law, vol. 5, no. 1, 2014, s. 53-59.

Yang, C. – Liao, H. 2010

Yang, Chyan – Liao, Hsien-Jyh: Using the robots.txt and Robots Meta Tags to Implement Online Copyright and a Related Amendment. Library Hi Tech, 28 (1), s. 94-106.

Virallislähteet

Hallituksen esitykset

HE 28/2004 vp. Hallituksen esitys Eduskunnalle laeiksi tekijänoikeuslain ja rikoslain 49 luvun muuttamisesta

HE 92/2018 vp. Hallituksen esitys eduskunnalle laiksi tekijänoikeuslain muuttamisesta

Euroopan unionin virallislähteet

Artificial Intelligence, Robotics, Privacy and Data Protection 2016

Artificial Intelligence, Robotics, Privacy and Data Protection: Room document for the 38th International Conference of Data Protection and Privacy Commissioners. October 2016.

Digitaalisten sisämarkkinoiden strategia Euroopalle, COM(2015)

Komission tiedonanto Euroopan parlamentille, neuvostolle, Euroopan talous- ja sosiaalikomitealle ja alueiden komitealle: Digitaalisten sisämarkkinoiden strategia Euroopalle. Bryssel 6.5.2015, COM(2015) 192 final.

WIPO

Draft Issues Paper on Intellectual Property Policy and Artificial Intelligence 2019

WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI), Second Session: Draft Issues Paper on Intellectual Property Policy and Artificial Intelligence, WIPO/IP/AI/2/GE/20/1, 2019.

Oikeustapakset

Euroopan unionin tuomioistuin

Infopaq I

C-5/08, Infopaq International A/S v Danske Dagblades Forening, 16.7.2009, ECLI:EU:C:2009:465

Infopaq II

C-302/10, Infopaq International A/S v Danske Dagblades Forening, 17.1.2012, ECLI:EU:C:2012:16

Premier League

Yhdistetyt asiat C-403/08 ja C-429/08, Football Association Premier League Ltd and Others v QC Leisure and Others and Karen Murphy v Media Protection Services Ltd, 4.10.2011, ECLI:EU:C:2011:798

Yhdysvallat

Bill Graham Archives v Dorling Kindersley Ltd.

Bill Graham Archives v Dorling Kindersley Ltd, 448 F. 3d 605 (2nd Cir. 2006)

Campbell

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994)

Google Books I

Authors Guild, Inc. v Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013)

Google Books II

Authors Guild, Inc. v. Google, Inc., 804 F.3d 202 (2nd Cir. 2015)

Harper & Row

Harper & Row v. Nation Enterprises, 471 U.S. 539 (1985)

HathiTrust

Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2nd Cir. 2014)

Oracle III

Oracle Am., Inc. v. Google, 886 F.3d 1179, 1179 (Fed. Cir. 2018)

Sony

Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984)

Stewart v. Abend

Stewart v. Abend, 495 U.S. 207 (1990)

Muut lähteet

Daddy's Car -kappale

Sony CSL: Daddy's Car: a song composed with Artificial Intelligence – in the style of the Beatles. <URL: https://www.youtube.com/watch?v=LSHZ_b05W7o>. Viitattu 11.3.2021.

Google Books Library Project

Google Books Library Project, <URL: <https://support.google.com/websearch/answer/9690276?hl=en>> Viitattu 10.8.2020.

"Data", Tieteen termipankki

Tieteen termipankin selitys sanalle "data". <URL: <https://tieteentermipankki.fi/wiki/Nimitys:data>>. Viitattu 15.3.2020.

"Informaatio", Tieteen termipankki

Tieteen termipankin selitys sanalle "informaatio". <URL: <https://www.tieteentermipankki.fi/wiki/Nimitys:informaatio>>. Viitattu 15.3.2020.

Lehdistötiedote, Google: Google Checks out Library Books

Lehdistötiedote, Google: Google Checks out Library Books. <URL: <http://googlepress.blogspot.com/2004/12/google-checks-out-library-books.html>> Viitattu 10.8.2020.

"Tieto", Tieteen termipankki

Tieteen termipankin selitys sanalle "tieto". <URL: <https://www.tieteentermipankki.fi/wiki/Nimitys:tieto>>. Viitattu 15.3.2020.

Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules

Max Planck Institute for Innovation and Competition: Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed

Modernisation of European Copyright Rules, Part B: Exceptions and Limitations, Chapter 1: Text and Data Mining (Article 3 COM(2016) 593), Version 1.1. <URL: https://www.ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/MPI_Position_Statement_Part_B_Chapter_1_Update23022017.pdf> Viitattu 26.6.2020.

The Next Rembrandt

The Next Rembrandt. <URL: <https://www.nextrembrandt.com/>>. Viitattu 11.3.2021.

The Verge 2016

The Verge 26.9.2016: “This AI-written pop song is almost certainly a dire warning for humanity”. <URL: <https://www.theverge.com/2016/9/26/13055938/ai-pop-song-daddys-car-sony>>. Viitattu 11.3.2021.

University of Toronto 2020

“U of T infectious disease expert’s AI firm now part of Canada’s COVID-19 arsenal”, 27.3.2020. <URL: <https://www.utoronto.ca/news/u-t-infectious-disease-expert-s-ai-firm-now-part-canada-s-covid-19-arsenal>>. Viitattu 15.12.2020.

Lyhenteet

Bernin sopimus

Bernin yleissopimus kirjallisten ja taiteellisten teosten suojaamisesta

DSM-direktiivi

Euroopan parlamentin ja neuvoston direktiivi (EU) 2019/790, annettu 17 päivänä huhtikuuta 2019, tekijänoikeudesta ja lähioikeuksista digitaalisilla sisämarkkinoilla ja direktiivien 96/9/EY ja 2001/29/EY muuttamisesta

DSM-strategia

Digitaalisten sisämarkkinoiden (Digital Single Market) strategia

Tietoyhteiskuntadirektiivi

Euroopan parlamentin ja neuvoston direktiivi 2001/29/EY, annettu 22 päivänä toukokuuta 2001, tekijänoikeuden ja lähioikeuksien tiettyjen piirteiden yhdenmukaistamisesta tietoyhteiskunnassa

Tietokantadirektiivi

Euroopan parlamentin ja neuvoston direktiivi 96/9/EY, annettu 11 päivänä maaliskuuta 1996, tietokantojen oikeudellisesta suojasta

Tietokoneohjelmadirektiivi

Euroopan parlamentin ja neuvoston direktiivi 2009/24/EY, annettu 23 päivänä huhtikuuta 2009, tietokoneohjelmien oikeudellisesta suojasta

TRIPS-sopimus

Sopimus tekijän- ja teollisoikeuksien kauppaan liittyvistä näkökohdista (Agreement on Trade-Related Aspects of Intellectual Property Rights)

WIPO

Maailman henkisen omaisuuden järjestö (World Intellectual Property Organization)

1. Johdanto

1.1 Johdatus aiheeseen

Konsulttiyhtiö IDC:n arvion mukaan maailmassa tulee olemaan 175 zetabittiä digitaalista tietoa vuoteen 2025 mennessä. Zetabitti vastaa yhteensä triljoonaa gigabittiä. Mikäli kaikki tämä digitaalinen tieto mahdutettaisiin DVD-levyihin, lopputuloksena syntyvä levypino kattaisi maapallon ja kuun välisen etäisyyden 23 kertaa.¹ Yksinään digitaalisen tiedon määrää ei kuitenkaan voida pitää itseisarvona datavetoisen yhteiskunnan kehittämisen kannalta. Tiedosta on kyettävä havainnollistamaan kaavoja ja syy-seuraussuhteita, jotta sen potentiaali voidaan maksimoida. Tässä työssä voidaan hyödyntää muun muassa tekstin- ja tiedonlouhintaa.²

Tekstin- ja tiedonlouhinta mahdollistaa kerätyn datan prosessoinnin, yhdistämisen ja jalostamisen uudeksi informaatioksi.³ Toisin sanoen, tekstin- ja tiedonlouhintaa voidaan kuvata prosessiksi, jossa yhdistetään eri lähteistä ja erilaisissa muodoissa kerättyä dataa toisiinsa liittäen kokonaisuuksiksi, joiden pohjalta on mahdollista saavuttaa uutta tietämystä ja oivalluksia. Tätä taustaa vasten ei ole yllättävää, että tekstin- ja tiedonlouhinnalla on suuri merkitys muun muassa tieteellisessä tutkimuksessa ja koneoppivien algoritmien kehittämisessä. Seuraavaksi esitellään joitakin tekstin- ja tiedonlouhinnan käytännön sovellutuksia.

Kanadalainen ohjelmistoyritys BlueDot antoi varoituksen maailmaa potentiaalisesti uhkaavasta viruksesta – joka tunnetaan nykyisin paremmin nimellä COVID-19 – jo ennen kuin kansalliset ja ylikansalliset terveysterveystoimet ehtivät asialle. Lisäksi yrityksen big data -analytiikkaan perustuva alusta onnistui ennustamaan viruksen alkuvaiheessa 11 Wuhanin tartunnoissa seuraavaksi seurannutta kaupunkia.⁴ BlueDotin kehittämän alustan viruksen havainnut teknologia perustuu koneoppimiseen ja luonnollisen kielen prosessointiin. Niiden pohjalta rakennettu tietokonealgoritmi on oppinut tunnistamaan yhteensä 150 erilaista taudinaiheuttajaa sille syötetystä datasta, joka koostuu tässä tapauksessa muun muassa uutisraporteista, lääketieteellisistä julkaisuista, ilmastoon ja

¹ Reinsel ym. 2018, s. 6-7.

² Huomautettakoon kuitenkin, että koneoppimisen standardimenetelmät tuottavat kausaalisten mallien sijaan tilastollisia malleja. Näin ollen menetelmillä ei useimmiten voida havaita syy-seuraussuhteita, koska menetelmät ovat puhtaasti tilastollisia. Ks. aihepiiristä tarkemmin esim.: Pearl 2019, s. 54-60.

³ Rosati 2019, s. 1-2.

⁴ University of Toronto 2020.

väestöön liittyvistä tiedoista sekä eläimillä leviäviä sairauksia koskevista raporteista. Algoritmin havaitessa epätavallisen tilanteen se käyttää lentoyhtiöiden lipunmyyntidataa ennustaakseen taudin leviämistä.⁵

Tekstin- ja tiedonlouhintaa käytetään enenevissä määrin apuna myös luovien töiden kehittämisessä. Musiikin osalta esimerkki tästä on Sonyn tutkijoiden johtama projekti, jonka lopputuloksena saatiin aikaan ensimmäinen tekoälyn kirjoittama pop-kappale. *Daddy's car* -nimisen teoksen⁶ luomistyössä hyödynnettiin Flow Machines -ohjelmistoa, jonka toiminta perustuu yli 10 000 eri kappaleen melodioita ja harmonioita sisältävän tietokannan analysointiin.⁷ Ohjelmisto laatii itse melodiansa, minkä jälkeen ihmisen panosta tarvitaan ainoastaan lyriikoiden kirjoittamiseen ja kappaleen tuottamiseen.⁸ Kyseessä olevassa tapauksessa kappaleen haluttiin muistuttavan The Beatles -yhtyeen tuotantoa, minkä vuoksi ohjelmiston analysoitavaksi syötettiin kyseiseen genreen kuuluvaa musiikkia.⁹ Lopputuloksesta voidaan olla musiikillisesti montaa mieltä, mutta on haastavaa kiistää, etteikö kehityssuunnalla tulisi olemaan vaikutuksensa musiikkiteollisuuden tulevaisuuteen.

Kuvataiteen saralla huomiota on puolestaan herättänyt *The Next Rembrandt* -projekti,¹⁰ jossa tekstin- ja tiedonlouhintaa hyödynnettiin Rembrandtin omakuvan luomiseen. Projekti aloitettiin kokoamalla laaja tietokanta taiteilijan maalauksista, minkä jälkeen niitä analysoitiin pikseli pikseliltä. Tämän jälkeen päätettiin maalauksen kohde ja suunniteltiin ohjelmisto, jonka tehtävänä oli ymmärtää Rembrandtin tyyliä ja luoda uusia piirteitä. Kun 2D-kuva teoksesta oli saatu valmiiksi, maalaus viimeisteltiin pyrkien matkimaan taiteilijan sivellinvetoja. Esimerkkejä vastaavista projekteista löytyy myös kirjallisuuden parista. Kaikkia mainittuja hankkeita yhdistää laajan datamassan syöttäminen koneoppivan ohjelmiston analysoitavaksi.

WIPO (World Intellectual Property Organization) julkaisi joulukuussa 2019 luonnospaperin,¹¹ joka toimii pohjustuksena immateriaalioikeuksien ja tekoälyn rajapinnoilla jäsenvaltioiden ja sidosryhmien kanssa käytävään keskusteluun. Luonnoksessa on hahmoteltu lähtökohtia aihepiirin ajankohtaisimpien kysymysten ratkaisemiseksi

⁵ University of Toronto 2020.

⁶ Daddy's Car -kappale.

⁷ The Verge 2016.

⁸ The Verge 2016.

⁹ The Verge 2016.

¹⁰ The Next Rembrandt.

¹¹ Draft Issues Paper on Intellectual Property Policy and Artificial Intelligence 2019.

käytäviin keskusteluihin. Yksi paikannetuista kysymyksenasetteluista koskee tekoälysovelluksen kehittämisessä käytettävän datan tekijänoikeuksia. Dataan saattaa nimittäin sisältyä tekijänoikeussuojattuja teoksia, joiden käyttäminen koneoppivan ohjelmiston opettamiseen voi olla tietyissä tilanteissa tulkittavissa tekijänoikeusloukkaukseksi. Tämä perustuu siihen lähtökohtaan, että tekstin- ja tiedonlouhinta sisältää useimmissa tapauksissa ainakin jonkinasteista kopiointia, joka taas voidaan tulkita kappaleen valmistamista koskevaa oikeutta loukkaavaksi toiminnaksi.¹² Oikeuskäytännössä omaksutun kannan mukaan tämä pätee myös silloin, kun louhinnan ja sitä kautta kopioinnin kohteena on vain lyhyt ote tekijänoikeudella suojatusta teoksesta.¹³ Kuvaillun skenaarion voidaan katsoa johtavan varsin haastavaan tilanteeseen sekä oikeudenhaltijoiden että dataa tekoälysovelluksensa kehittämiseksi louhivan tahon kannalta, sillä koneoppivan ohjelmiston opettamiseen tarvitun datan määrä on tyypillisesti varsin suuri.

Tilanteeseen on reagoitu parhaillaan implementointivaiheessa olevassa EU:n DSM-direktiivissä¹⁴. Direktiivin 3 artiklassa on säädetty tekijänoikeuden poikkeuksesta tai rajoituksesta koskien tekstin- ja tiedonlouhintaa tieteellistä tutkimusta varten. Direktiivin 4 artiklassa puolestaan säädetään vastaavan tyyppisestä tekijänoikeuden poikkeuksesta tai rajoituksesta, mutta kohdistuen tieteellisen tutkimuksen sijaan muihin tekstin- ja tiedonlouhinnan taustalla oleviin tarkoituksiin. Kärjistäen voisikin jo tässä vaiheessa todeta, että 3 artiklassa tarkoitettu tekstin- ja tiedonlouhinta perustuu ei-kaupalliseen tarkoitukseen, kun taas 4 artiklan mukaisen tekstin- ja tiedonlouhinnan taustalla on kaupallisia tarkoituksia.¹⁵ Artiklojen eroavaisuudet eivät kuitenkaan pääty niiden taustalla olevien tarkoituksien päinvastaisuuteen. 4 artiklan 3 kohdassa nimittäin annetaan oikeudenhaltijalle mahdollisuus yksipuolisesti pidättää oikeutensa artiklan 1 kohdassa tarkoitettuun teosten ja muun suojatun aineiston käyttöön.¹⁶ 3 artiklassa oikeudenhaltijoille ei ole myönnetty tällaista oikeutta.

¹² Geiger ym. 2018 s. 6.

¹³ C-5/08 (Infopaq), kohdat 54 ja 55.

¹⁴ Euroopan parlamentin ja neuvoston direktiivi (EU) 2019/790, annettu 17 päivänä huhtikuuta 2019. tekijänoikeudesta ja lähioikeuksista digitaalisilla sisämarkkinoilla ja direktiivien 96/9/EY ja 2001/29/EY muuttamisesta.

¹⁵ 4 artiklan tarkoittamasta tekstin- ja tiedonlouhinnasta käytetään tässä tutkielmassa pääasiallisesti nimityksiä ”kaupallisessa tarkoituksessa toteutettava tekstin- ja tiedonlouhinta” tai ”kaupallinen tekstin- ja tiedonlouhinta”.

¹⁶ 4 artiklan 3 kohdan oikeudenhaltijoille suomasta mahdollisuudesta käytetään jäljempänä tutkielmassa myös nimitystä ”opt out -mekanismi”.

DSM-direktiivin 4 artiklan 3 kohdassa tehty lainsäädäntöratkaisu perustuu nähdäkseni tarpeeseen tasapainottaa yksityistä ja yleistä etua tekijänoikeudellisissa kysymyksissä. Taustoittaaksemme ajatusta on syytä johdattaa lukija lyhyesti myös tähän tekijänoikeuden perimmäisiä tavoitteita koskevaan aihepiiriin. Historiallisesta perspektiivistä katsottuna tekijänoikeuslainsäädännön tärkeimmäksi tavoitteeksi on nostettu kirjallisuuden ja taiteen edistäminen.¹⁷ Tavoite heijastuu luonnollisesti myös lainvalmisteluaineistossa, jossa yhtäältä painotetaan tekijänoikeuden roolia yhteiskunnallisena kannustimena luovalle toiminnalle ja toisaalta tunnistetaan tarve huomioida myös muut kuin oikeudenhaltijoiden intressit.¹⁸

Teknologinen kehitys on kuitenkin osaltaan vahvistanut vastakkainasettelua oikeudenhaltijoiden suojaamisen ja teosten hyödyntäjien intressien välille. *Haarmannin* mukaan vastakkainasettelun luomien jännitteiden purkaminen edellyttää lainsäätäjältä ”selkeän kokonaisnäkemyksen” löytämistä.¹⁹ Tietyissä tilanteissa eri intressiryhmien etujen ja oikeuksien välinen epätasapaino voinee johtaa teknologisen kehityksen ja jossain määrin myös siihen liittyvien luovien ratkaisuiden kangistumiseen. Toisaalta myöskään teknologisen kehityksen vauhdittamista ei voida pitää jonkinlaisena yhteiskuntaa läpileikkaavana arvoperustana, jonka pohjalta oikeudenhaltijoiden etujen rajoittamista voitaisiin perustella miltei loputtomiin.

Tekijänoikeuslainsäädännössä on ajan mittaan kehitetty ratkaisuja yleisen ja yksityisen edun tasapainottelemista koskevaan problematiikkaan. Tekijänoikeudelle on asetettu nimenomaisia rajoituksia ja poikkeuksia, joiden puitteissa tietty tekijänoikeussuojan piiriin kuuluva toiminta voidaan sallia. Rajoitukset ja poikkeukset mahdollistavatkin teosten laillisen hyödyntämisen tavalla, joka ilman rajoitusten ja poikkeusten olemassaoloa tulkittaisiin tekijänoikeusloukkaukseksi. Rajoitusten ja poikkeusten luomista on perusteltu muun muassa yksityiseen etuun ja perusoikeuksiin sekä sivistyksellisiin ja sosiaalisiin syihin vedoten.²⁰ Rajoituksia on kuitenkin tulkittava suppeasti,²¹ minkä voitaneen arvioida johtuvan pyrkimyksestä varoa liiallista puuttumista tekijänoikeuslainsäädännön oikeudenhaltijoiden etua suojelevaan funktioon.

¹⁷ Haarmann 2005, s. 9.

¹⁸ Ks. esim. HE 28/2004 vp. s.7 ja HE 92/2018 vp. s. 3.

¹⁹ Haarmann 2014, s. 46.

²⁰ Haarmann 2014, s. 94.

²¹ C-5/08 (Infopaq), kohta 56.

Tekijänoikeuden ja tekoälykehityksen rajapinnassa käydyssä keskustelussa on kiinnitetty paljon huomiota siihen, missä määrin tekoälyn luomat teokset voivat nauttia tekijänoikeussuojaa. Vastaavaa immateriaalioikeussuojan myöntämisedellytyksiin liittyvää diskurssia on käyty myös patenttioikeuden saralla. Tekijänoikeuden ja tekstin- ja tiedonlouhinnan välisen suhteen ei voida sanoa jääneen unholaan, mutta sitä koskevan analyysin tarpeen voitaneen katsoa korostuneen tilanteessa, jossa EU on uudella lainsäädännöllään ilmaissut kantansa aihekokonaisuuteen. WIPO:n luonnospaperiin saadun lausuntopalautteen perusteella on pääteltävissä, että eri oikeusjärjestysten lähestymistavat tekstin- ja tiedonlouhinnan ja tekijänoikeuden väliseen problematiikkaan eroavat jossain määrin toisistaan erityisesti puhuttaessa kaupallisessa tarkoituksessa toteutetusta louhinnasta.

Mahdollisuus analysoida suuria ja mahdollisimman laadukkaita datamassoja on tärkeää koneoppimisen ja tulevaisuuden tekoälykehityksen kannalta. Ei ole minkään talousalueen intressien mukaista jäädä jälkeen tekoälyn hyödyntämiseen perustuvassa kilpajuoksussa. Tämän vuoksi tekstin- ja tiedonlouhintaan liittyvä sääntely yhtenä tekoälyn laaja-alaisen hyödyntämisen kulmakivenä on ajankohtainen kokonaisuus tutkittavaksi.

1.2 Tutkimuskohteet ja -kysymykset

Tutkielman yleisenä viitekehyksenä on tekijänoikeussäätelyn ja tekstin- ja tiedonlouhinnan välinen suhde. Suhdetta tarkastellaan ensisijaisesti EU-lainsäädäntöön perustuen, mutta paremman kokonaiskuvan saamiseksi myös oikeusvertailun kautta. EU-säätelyn osalta tutkielmassa keskitytään sekä säädöksiin, jotka olivat relevantteja ennen DSM-direktiivin laatimista että itse DSM-direktiiviin. Ensiksi mainitun säädösjoukon muodostavat tietoyhteiskuntadirektiivi²², tietokantadirektiivi²³ ja tietokoneohjelmadirektiivi²⁴. Näiden säädösten osalta tutkielmassa suunnataan katse erityisesti kappaleen valmistamista koskevaan oikeuteen liittyviin säännöksiin, sillä DSM-direktiivin tutkielman kannalta relevantit poikkeukset ja rajoitukset kohdistuvat juuri kappaleen valmistamista koskevaan yksinoikeuteen. Näin toimien pyritään muodostamaan paras mahdollinen kuva kaupallisen tekstin- ja tiedonlouhinnan tekijänoikeudellisen aseman muutoksesta DSM-direktiivin seurauksena. Kun huomio keskitetään nimenomaan kaupalliseen tekstin- ja tiedonlouhintaan,

²² Euroopan parlamentin ja neuvoston direktiivi 2001/29/EY, annettu 22 päivänä toukokuuta 2001, tekijänoikeuden ja lähioikeuksien tiettyjen piirteiden yhdenmukaistamisesta tietoyhteiskunnassa.

²³ Euroopan parlamentin ja neuvoston direktiivi 96/9/EY, annettu 11 päivänä maaliskuuta 1996, tietokantojen oikeudellisesta suojasta.

²⁴ Euroopan parlamentin ja neuvoston direktiivi 2009/24/EY, annettu 23 päivänä huhtikuuta 2009, tietokoneohjelmien oikeudellisesta suojasta.

on luonnollista, että DSM-direktiivin artikloista relevanteimmaksi nousee jo edellisessä alaluvussa lyhyesti esitelty 4 artikla.

DSM-direktiivin on määrä tulla implementoiduksi osaksi jäsenvaltioiden kansallista lainsäädäntöä kesäkuun 7. päivään mennessä vuonna 2021. Direktiivin 4 artiklassa säädetään seuraavasti:

Jäsenvaltioiden on säädettävä poikkeuksesta tai rajoituksesta direktiivin 96/9/EY 5 artiklan a alakohdassa ja 7 artiklan 1 kohdassa, direktiivin 2001/29/EY 2 artiklassa, direktiivin 2009/24/EY 4 artiklan 1 kohdan a ja b alakohdassa sekä tämän direktiivin 15 artiklan 1 kohdassa säädettyihin oikeuksiin, kun kyseessä on kappaleen valmistaminen ja kopiointi tekstin- ja tiedonlouhintaa varten teoksista ja muusta suojatusta aineistosta, johon on laillinen pääsy;

Edellä olevan 1 kohdan mukaisesti valmistettuja kappaleita ja kopioita voidaan säilyttää niin kauan, kuin tekstin- ja tiedonlouhintaa varten on tarpeen;

Edellä 1 kohdassa säädettyä poikkeusta tai rajoitusta sovelletaan edellyttäen, että oikeudenhaltijat eivät ole nimenomaisesti pidättäneet oikeuksiaan kyseisessä kohdassa tarkoitettuun teosten ja muun suojatun aineiston käyttöön asianmukaisella tavalla, kuten koneluettavilla keinoilla, kun kyseessä on verkossa yleisön saataviin saatettu sisältö;

Tämä artikla ei vaikuta tämän direktiivin 3 artiklan soveltamiseen.

Direktiivissä säädetty poikkeus tai rajoitus kohdistuu näin ollen tietokantasuojaa, kappaleen valmistamista koskevaa oikeutta, tietokoneohjelmien tekijänoikeudellista suojaamista ja lehtijulkaisujen suojaa verkkokäytössä koskeviin säännöksiin. Artiklan tarkoituksena on mahdollistaa tekstin- ja tiedonlouhinta myös muussa kuin tieteelliseen tutkimukseen perustuvassa käyttötarkoituksessa.

Tutkielmassa tullaan hyödyntämään myös oikeusvertailevaa tutkimusmetodia.²⁵ Oikeusvertailun osalta tutkimuskohteena on DSM-direktiivin 4 artiklan lisäksi Yhdysvaltojen tekijänoikeusjärjestelmän ytimeen sijoittuva fair use -säännös sekä

²⁵ Tutkielmassa hyödynnettyjä tutkimusmenetelmiä esitellään kokonaisvaltaisemmin luvussa 1.3.

säännöksen analysointi tekstin- ja tiedonlouhinnan tekijänoikeudellisen statuksen kannalta relevanteissa yhdysvaltalaisissa oikeustapauksissa.

Edellä kuvatut tutkimuskohteet jäisivät ilman tutkimuskysymysten asettamista hajanaisiksi pisteiksi, jotka eivät oikein yhdistyisi toisiinsa. Tutkimuskysymysten valinnassa on painotettu selkeyttä ja johdonmukaisuutta, jotta tutkimuskohteet saataisiin mahdollisimman hyvin raamitettua OTM-tutkielman puitteisiin. Tutkimuskysymykset ovat seuraavat:

- 1. Miten kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan ja tekijänoikeuden välinen suhde muuttuu DSM-direktiivin myötä EU:ssa?*
- 2. Miten EU:n muuttunut oikeustila vertautuu tekstin- ja tiedonlouhinnan ja tekijänoikeuden väliseen suhteeseen Yhdysvalloissa?*

Tutkielma on rajattu edellä esitetyllä tavalla koskemaan DSM-direktiivin 4 artiklaa. Näin ollen esimerkiksi saman direktiivin 3 artiklan mukaista tekstin- ja tiedonlouhintaa tieteellistä tutkimusta varten ei käsitellä kovinkaan syvällisesti. 3 artikla tarjoaa kuitenkin erilaisen muotoilunsa kautta mielenkiintoisen vertailukohdan 4 artiklaan, minkä vuoksi sitä tullaan hyödyntämään paikoin tästä perspektiivistä. 4 artiklan 1 kohdassa mainittuja muita säännöksiä käsitellään tutkielmassa niin ikään siinä määrin kuin se koetaan hyödylliseksi tutkimuskysymyksiin vastaamisen kannalta. DSM-direktiivin kesään 2021 ulottuvasta implementointiajasta johtuen tutkielmassa ei tulla analysoimaan yksittäisten jäsenvaltioiden tekemiä implementointiratkaisuja, sillä kyse olisi tutkielman aikataulusta johtuen tällöin varsin spekulatiivisesta tutkimusotteesta. Etenkin Suomessa jo nyt toteutetusta implementointityöstä on kuitenkin löydettävissä melko paljon materiaalia, jota hyödynnetään ensisijaisesti ajatuksia herättävässä tarkoituksessa.

1.3 Tutkimusmenetelmä ja tutkimuksen rakenne

Tutkimusmenetelmät eli metodit kuuluvat olennaisena osana oikeustieteen harjoittajan työkalupakkiin. Silloinkin, kun tutkimusta tehdään kiinnittämättä erityistä huomiota käytettäviin metodeihin, niihin joudutaan ennemmin tai myöhemmin tukeutumaan työtä edistettäessä.²⁶ Oikeustieteessä metodivalinnat perustuvat pitkälti valittuun tutkimusongelmaan, jota puolestaan täsmennetään huolellisesti valituilla

²⁶ Hirvonen 2011, s. 6.

tutkimuskysymyksillä.²⁷ Käsillä olevan tutkielman tutkimuskysymysten perustuessa yhtäältä voimassa olevan oikeuden tulkintaan sitä edeltäneen oikeustilan kautta ja toisaalta sen vertailemiseen toiseen oikeusjärjestykseen nähden, koen metodisen pluralismin perustelluksi tutkielman kannalta.

Määttä on pohtinut metodista avoimuutta ja pluralismia ympäristöoikeudessa katsoen, että metodisen vastakkainasettelun sijaan etualalle tulisi asettaa erilaisten tutkimuksellisten lähestymistapojen toisiaan tukeva vaikutus.²⁸ Myös *Husa* on katsonut, että oikeustieteellistä tutkielmaa laatiessa voidaan yhdistää eri metodeja tai liikkua niiden välimaastossa.²⁹ Nähdäkseni eri metodien yhdistelemistä koskeva traditio olisi ulotettavissa monelle muullekin sellaiselle oikeudenalalle, jonka piiriin sijoittuvan tutkimuksen tuloksellisuus vaatii tutkimuskohteen kokonaisvaltaista käsittelyä. OTM-tutkielman osalta tulee kuitenkin huomioida, että tutkielman suhteellisen rajattu enimmäispituus saattaa asettua esteeksi kovin runsaalle metodien limittäiselle käyttämiselle.

Tutkielman pääasialliseksi metodiksi on valittu lainoppi, joka on vastaavassa asemassa myös suuressa osassa muista oikeustieteen alan tutkimustöistä. Lainopin tutkimuskohtena on voimassa oleva oikeus, eli toisin sanoen se, mitä tietyssä oikeusjärjestyksessä voimassaolevat oikeusnormit pitävät sisällään. Lainopin tehtäväksi on määritelty normikannanottojen ja tulkintakannanottojen esittäminen.³⁰ Normikannanottojen kautta systematisoidaan oikeusnormeja osaksi voimassaolevaa oikeutta, kun taas tulkintakannanotot pyrkivät esittämään väitteitä oikeusnormin sisällöstä. Tätä rajanvetoa on kuitenkin kritisoitu sillä perusteella, että kumpaankin kannanottotapaan sisältyy myös toiseen tapaan liittyviä väitteitä.³¹ Tutkielmassa otetaan kritiikki huomioon, mutta pitäydytään siitä huolimatta jaottelussa, jossa voimassaolevaa oikeutta pyritään yhtäältä systematisoimaan ja toisaalta tulkitsemaan lainopillisen metodin avulla. Tutkielman aiheen kannalta tämä tarkoittaa sitä, että DSM-direktiivin 4 artiklaa ja muita huomioitavia säännöksiä asemoidaan ensinnä osaksi vallitsevaa oikeusjärjestystä ja sen jälkeen tulkitaan niiden merkityssisältöä.

²⁷ Kolehmainen 2015, s. 3 ja 6.

²⁸ Määttä 2015, s. 59.

²⁹ Husa ym. 2008, s. 19.

³⁰ Hirvonen 2011, s. 22.

³¹ Hirvonen 2011, s. 22.

Tutkielmassa täydennetään lainopillista metodologia osittain oikeuspoliittisella tutkimuksella, jota kutsutaan *de lege ferenda* -tutkimukseksi. De lege ferenda -tutkimuksessa voimassaolevien oikeusnormien systematisointiin ja tulkintaan keskittyvän tavanomaisen lainopin sivutuotteena voidaan arvioida tulevan lainsäädännön perusteiksi sopivia sääntelymalleja.³² De lege ferenda -tulkinnat antavat tutkijalle vapauden analysoida mahdollisia lainsäädäntömalleja esimerkiksi yhteiskunnallisen tarkoituksenmukaisuusharkinnan kautta.³³

Tutkielmassa arvioidaan muun muassa DSM-direktiivin 4 artiklan soveltuvuutta ottaa riittävällä tavalla huomioon tekijänoikeudellinen yleisen ja yksityisen edun välinen tasapainottelu sekä artiklan yhteensopivuutta yhteiskunnan digitalisoitumisen kanssa. Molemmilta osin de lege ferenda -tutkimuksella on tärkeä perinteistä lainoppia täydentävä merkitys. Perinteisen lainopin ja de lege ferenda -tutkimuksen rajapinnassa on syytä huomioda myös se, että DSM-direktiiviä ei ole vielä implementoitu osaksi jäsenvaltioiden omaa lainsäädäntöä. Näin ollen oikeuspoliittinen tutkimus voitaisiin kohdistaa myös siihen, millaisia lainsäädäntömalleja jäsenvaltioilla olisi mahdollista toteuttaa direktiiviä implementoidessaan. Direktiivin 4 artikla jättää kuitenkin melko vähän mahdollisuuksia sen sisällöstä poikkeaville kansallisille ratkaisuille, minkä vuoksi koen de lege ferenda -tutkimuksen hyödyllisemmäksi siinä kontekstissa, millaisia sääntelyllisiä vaihtoehtoja ylipäänsä direktiivin mukaiselle sisällölle olisi ollut tai tulee olemaan yhteiskunnallisen kehityksen kannalta.

Tutkielman jälkimmäisen tutkimuskysymyksen avulla tarkastellaan vierasta oikeutta ja siitä paikannettavia vaihtoehtoisia ratkaisumalleja tekstin- ja tiedonlouhinnan ja tekijänoikeussääntelyn välisen suhteen problematiikkaan. Näin ollen tutkielmassa hyödynnetään ainakin termiä löyhästi tulkiten myös vertailevaa oikeustieteen metodologia. Vertailevan näkökulman hyödyntäminen perustuu tutkielmassa kuitenkin ensi sijassa pyrkimykseen rikastaa lainopillista tutkimusta vieraan oikeuden sääntelymalleilla ja sitä kautta tuoda uusia näkökulmia tutkimuskohteen käsittelyyn.³⁴ Husa katsoo, että oikeusvertailulle tyypillisesti asetetut metodiset vaatimukset eivät ole kuvatussa käyttötarkoituksessa yhtä kireitä.³⁵ Tutkielmassa pyritään tästä huolimatta tutkimaan vieraan

³² Kolehmainen 2015, s. 3.

³³ Kolehmainen 2015, s. 3.

³⁴ Oikeusvertailevan metodin lainoppia rikastavasta merkityksestä ks. esim. Husa 2010, s. 713; Ervo 2008, s. 320.

³⁵ Husa 2010, s. 713.

oikeuden lähestymistapoja tekstin- ja tiedonlouhinnan tekijänoikeudelliseen sääntelyyn mahdollisimman syvällisesti kulloinkin kyseessä olevan oikeusjärjestyksen perspektiivistä. Vain tällä tavoin toimimalla vaihtoehtoiset toimintamallit voidaan esittää rinnasteisina eurooppalaiseen sääntelytapaan nähden. Pääasiallisen tutkimuskohteen kietoutuessa ensi sijassa EU-alueen sääntelyn ympärille vieraan oikeuden vertailukohtat haetaan luonnollisesti Euroopan ulkopuolelta, eli tässä tapauksessa Yhdysvalloista.

Tutkielman rakenne perustuu portaittaiseen etenemistapaan, jossa lukija perehdytetään aluksi tekstin- ja tiedonlouhintaan prosessina sekä tekijänoikeuden merkitykseen prosessin toteuttamisen kannalta. Nämä lähtökohdat muodostavat tutkielman toisen luvun. Kolmannessa luvussa esitellään tekstin- ja tiedonlouhinnan kannalta relevantit tietoyhteiskuntadirektiivin, tietokantadirektiivin, tietokoneohjelmadirektiivin ja DSM-direktiivin säännökset. Lisäksi pohditaan, mitä muutoksia DSM-direktiivi tuo tekstin- ja tiedonlouhinnan ja tekijänoikeussääntelyn väliseen suhteeseen EU:ssa. Neljännessä luvussa keskitytään puolestaan analysoimaan Yhdysvaltojen fair use -säännökseen perustuvaa tekijänoikeudellista traditiota ja siihen liittyviä oikeustapauksia suhteessa tekstin- ja tiedonlouhintaan. Analyysin pohjalta luvussa vertaillaan Yhdysvaltojen relevanttia sääntely-ympäristöä DSM-direktiivin muuttamaan eurooppalaiseen oikeustilaan. Tutkielman viidennessä luvussa analysoidaan esiteltyjä sääntelymalleja tarkemmin ja esitetään johtopäätöksiä. Johtopäätösten tarkoituksena ei ole valita absoluuttisesti parasta lainsäädäntömallia, vaan pohtia molempien hyviä ja huonoja puolia. Johtopäätösten pääasiallinen tavoite onkin nimenomaan tutkielman tutkimuskysymyksiin vastaaminen.

2. Tekstin- ja tiedonlouhinnasta

2.1 Mitä on tekstin- ja tiedonlouhinta?

Jotta tekstin- ja tiedonlouhintaa voidaan käsitellä tekijänoikeudellisesta perspektiivistä, on ensiksi selvennettävä, mitä tekstin- ja tiedonlouhinnalla tarkoitetaan ja mitä vaiheita louhinnan sisällyttävään prosessiin tyypillisesti lukeutuu. Tämän tutkielman puitteissa ei kuitenkaan ole syytä pureutua näihin vaiheisiin erityisen syvällisesti, vaan huomiota kiinnitetään ensisijaisesti niiden tekijänoikeudelliseen relevanttiuteen. Toisin sanoen, käsillä olevan luvun tarkoituksena on muodostaa lukijalle myöhemmän tekijänoikeudellisen analyysin ymmärtämiseksi tarvittava substanssipohja tekstin- ja tiedonlouhinnasta.

DSM-direktiivin 2 artiklan 2 kohdassa tekstin- ja tiedonlouhinnan määritellään tarkoittavan ”automaattista analyysitekniikkaa, jonka tarkoituksena on analysoida digitaalisessa muodossa olevaa tekstiä ja dataa tietojen, esimerkiksi muttei yksinomaan mallien, suuntausten tai korrelaatioiden, tuottamiseksi”. Edelleen direktiivin johdanto-osan 8 perustelukappaleessa todetaan seuraavasti: ”Uusi tekniikka mahdollistaa digitaalisessa muodossa olevan tiedon, kuten tekstin, äänen, kuvien tai datan, automaattisen laskennallisen analyysin, joka tunnetaan yleisesti tekstin- ja tiedonlouhintana. Tekstin- ja tiedonlouhinta mahdollistaa suurien tietomäärien käsittelemisen uuden tietämyksen hankkimiseksi ja uusien suuntausten havaitsemiseksi.”

DSM-direktiivin johdanto-osan 8 perustelukappaleen sanamuodot ovat nähdäkseni hieman ristiriitaisia: yhtäältä puhutaan digitaalisessa muodossa olevan tiedon automaattisesta laskennallisesta analyysistä, jossa teksti mainitaan vain yhtenä kyseisen tiedon muodoista, kun taas prosessista kokonaisuudessaan käytetään nimitystä tekstin- ja tiedonlouhinta. Johdonmukaisempi kattotermi kuvaamaan direktiivissä tarkoitettua toimintaa olisi ollut kaikessa yksinkertaisuudessaan tiedonlouhinta. Perustelen näkökulmaa sillä, että tekstinlouhinta voidaan lukea yhdeksi tiedonlouhinnan alakategoriaksi.

Digitaalista tietoa syntyy nyky-yhteiskunnassa valtavia määriä, eikä sitä voida luonnehtia missään määrin homogeeniseksi joukoksi. Erilaisia louhinnan kohteena olevia tietomuotoja voivat olla tekstidatan lisäksi muun muassa ääni ja videot.³⁶ Näin ollen tiedonlouhinnalla voidaan tutkielman viitekehyksessä tarkoittaa minkä tahansa tietomuodon louhintaa, kun

³⁶ Aggarwal 2015, s. 2.

taas tekstinlouhinnan merkityssisältö rajoittuu ainoastaan kyseisessä tietomuodossa olevan aineiston louhintaan. DSM-direktiivin sanavalintojen vuoksi tutkielmassa pyritään kuitenkin käyttämään säännönmukaisesti tekstin- ja tiedonlouhinnan käsitettä aina, kun ei ole tarkoituksena käsitellä tietyssä tietomuodossa olevan datan louhintaa erityisesti.

DSM-direktiivin määrittely tekstin- ja tiedonlouhinnan tarkoittamalle analyysitekniikalle on puolestaan melko tyydyttävä. Tekstin- ja tiedonlouhintaa voi kuitenkin hieman kontekstista riippuen kuvailla monin eri tavoin. Louhintaa näkee terminä käytettävän myös täysin muissa yhteyksissä, kuten esimerkiksi kaivostoimintaan liittyen. Tällöin tavoitteena saattaa olla esimerkiksi kullan louhiminen massiivisen kivimassan seasta. Prosessi ei taustaidealtaan juurikaan eroa tekstin- ja tiedonlouhinnasta: siinä missä kultaa louhiessa pyritään löytämään edes pieni määrä kultaa kivimassan keskeltä, tekstin- ja tiedonlouhinnassa tarkoituksena on löytää uutta informaatiota ja tehdä siihen pohjautuvia havaintoja suuren datamäärän seasta.³⁷

Havainnollistavana esimerkkinä voidaan käyttää elintarvikekauppaketjujen kanta-asiakasjärjestelmiä. Kanta-asiakasjärjestelmän voi katsoa perustuvan vaihtokauppaan, jossa asiakkaat hyötyvät ostoksistaan kertyvistä pisteistä, joita he voivat käyttää tulevaisuudessa, kun taas kauppaketju saa tietoonsa korttia käyttävien asiakkaiden tekemät ostokset. Ostokset eivät isona ja sekavana ryppäänä merkitse välttämättä mitään, mutta oikeaa tiedonlouhintatekniikkaa soveltamalla kauppaketjun on mahdollista saada selville esimerkiksi erilaisia korrelaatioita siitä, minkälaisia muita tuotteita tiettyä yksittäistä tuotetta ostavat henkilöt hankkivat.

DSM-direktiivin mainitessa tekstinlouhinnan otsikkotasolla rinnasteisena tiedonlouhinnan kanssa lienee syytä esitellä tämän yksittäisen tietomuodon louhinnan pääseikkoja lyhyesti. Tekstidatan määrä on kasvanut voimakkaasti viime vuosina, minkä seurauksena myös sen louhintaan käytettävä teknologia on ottanut edistysaskelia.³⁸ Tekstinlouhinta on poikkitieteellinen ala, jonka piiriin kuuluu tutkimusyhteisöjä muun muassa tiedonlouhinnan, luonnollisen kielen prosessoinnin, tiedonhaun ja koneoppimisen saroilta.³⁹ Tekstinlouhinta ei pääpiirteiltään eroa muiden tietomuotojen louhinnasta, sillä tekstinlouhinnassa pyritään laskennallisten prosessien avulla muuttamaan rakenteettomia sähköisiä tekstejä rakenteisiksi ja käyttämään tilastollisia metodeja uuden informaation ja uusien kaavojen löytämiseksi

³⁷ Han – Kamber – Pei 2011, s. 5-6.

³⁸ Aggarwal 2012, s. 1.

³⁹ Aggarwal 2012, s. 8-9.

prosessoidusta datasta.⁴⁰ Tällaista informaatiota voivat olla tekstinlouhinnan saralla esimerkiksi suositut kirjallisuuden aihepiirit viimeisten kymmenen vuoden ajalta tai asiakastyytyväisyyden arviointia ja ymmärtämistä käyttäjien tuotteista kirjoittamien kommenttien perusteella.⁴¹

Tekstin- ja tiedonlouhinta tarvitsee polttoaineekseen dataa. Nykypäivänä käytännössä kaikki automatisoidut järjestelmät tuottavat jonkin tyyppistä dataa joko vianmäärittelyksellisiin tai analyttisiin tarkoituksiin.⁴² Suurin osa jokapäiväisistä transaktioista, kuten pankkiautomaatin käyttäminen tai luottokortilla maksaminen tuottavat automatisoidusti dataa. Sama pätee esineiden internetin aikakaudella myös esimerkiksi teollisuuskoneisiin, joiden toimintaa tarkkaillaan sensoreiden avulla tai ylipäänsä älykkäisiin laitteisiin, jotka kykenevät kommunikoimaan keskenään.⁴³ Tekijänoikeuden kontekstissa tällainen data ei ole kuitenkaan järkevää kiinnostavaa puhtaasti siitä syystä, että sen ei ole mahdollista nauttia tekijänoikeussuojaa. Myös tekijänoikeudella suojattavissa olevaan materiaaliin, kuten kirjallisuuteen, musiikkiin ja taiteeseen sisältyy kuitenkin suurissa määrin dataa, jonka louhiminen on hyödyllistä. Juuri tällaista dataa louhittaessa saavutaan rajanvetopisteeseen, joka on tämän tutkielman ytimessä. Tällöin käydään kuvainnollisesti köydenvetoa siitä, missä määrin ja millä ehdoilla tekijänoikeudella suojattua aineistoa voidaan käyttää tekstin- ja tiedonlouhinnan polttoaineena prosesseissa, joiden lopputuloksena voimme tehdä uusia havaintoja meitä ympäröivästä yhteiskunnasta.

DSM-direktiivin sanavalinnoissa termit data ja tieto voidaan tulkita merkityssisällöiltään sekä toisiaan vastaaviksi että hieman toisistaan eroavaisiksi. Havainto perustuu siihen, että louhittava aineisto määritellään jo otsikkotasolla tiedoksi (”tekstin- ja tiedonlouhinta”), mutta direktiivin 2 artiklan 2 alakohdassa aineistosta puhutaan myös datana, josta voidaan louhinnan kautta tuottaa tietoa. Tutkielman terminologisen selkeyden vuoksi onkin syytä linjata, mikä merkityssisältö termeille annetaan tässä yhteydessä.

Tieteen termipankissa data määritellään digitaalseksi aineistoksi, johon ei liity tulkintaa.⁴⁴ Tiedolla tarkoitetaan puolestaan ymmärrettyä ja sisäistettyä informaatiota, kun taas datan ja tiedon välimaastoon sijoittuvaa informaatiota kuvataan rakenteeksi, jonka ymmärtäminen

⁴⁰ Sag 2019, s. 294-295.

⁴¹ Han – Kamber – Pei 2011, s. 14.

⁴² Aggarwal 2015, s. 1.

⁴³ Aggarwal 2015, s. 1.

⁴⁴ ”Data”, Tieteen termipankki.

antaa tietoa jostakin asiasta.⁴⁵ Tutkielman koherenssi ja käsitteellinen yhdenmukaisuus DSM-direktiivin kanssa edellyttävät kuitenkin nähdäkseni sitä, että tietoa ja dataa pidetään pitkälti synonyymeina keskenään. Tällöin tilanteissa, joissa halutaan keskittyä louhinnan tuotteena syntyvään ymmärrettyyn ja sisäistettyyn tietoon korostetaan tätä valintaa erikseen siitä mainitsemalla tai käyttämällä esimerkiksi termiä ”uusi tieto”. Louhinnan kohteena olevasta aineistosta voidaan tällöin puhua sekä datana että tietona ilman vaaraa väärinymmärryksistä.

2.2 Tekstin- ja tiedonlouhintaprosessin pääpiirteet

Louhimiskelpoisten tietomuotojen suuresta määrästä ja monista muista tekijöistä, kuten valitusta tiedonlouhintatekniikasta, johtuen tekstin- ja tiedonlouhintaprosessien vaiheet ovat varsin yksilöllisiä. Tutkielman tavoitteiden ja tutkimuskysymyksiin vastaamisen kannalta ei ole tarkoituksenmukaista sukeltaa liian syvälle näihin variaatioihin, vaan keskittää huomio tiedonlouhintaprosessin yleisiin vaiheisiin. Seuraavaksi perehdytäänkin lyhyesti tekstin- ja tiedonlouhintaprosessin pääpiirteisiin. Pyrkimyksenä on lisäksi kytkeä piirteitä tekijänoikeuteen niiltä osin kuin se on kulloinkin mahdollista tai tarkoituksenmukaista. Huomautettakoon kuitenkin, että tekstin- ja tiedonlouhinnan tarkempi tekijänoikeudellinen analyysi on toteutettu tutkielman myöhemmissä luvuissa, minkä vuoksi tässä yhteydessä pitäydytään varsin yleistasoisessa kuvailussa.

Aggarwalin mukaan tiedonlouhintaprosessi koostuu tyypillisesti kolmesta pääasiallisesta vaiheesta, jotka ovat datan kerääminen, datan esikäsittely ja datan analysointi.⁴⁶ *Rosatin* on tekstin- ja tiedonlouhinnan tekijänoikeudellista relevanssia käsitellessään myöskin eritellyt kolme tekstin- ja tiedonlouhintaprosesseissa yleisesti toistuvaa vaihetta, jotka ovat pääsy sisältöön (access to content), sisällön eriyttäminen ja/tai kopiointi (extraction and/or copying of content) sekä tekstin ja/tai datan louhiminen ja tietämyksen lisääminen (mining of text and/or data and knowledge discovery).⁴⁷

Edellä mainitut tulkinnat tyypillisen tiedonlouhintaprosessin vaiheista täydentävät hyvin toisiaan. Tämä johtuu siitä, että *Rosatin* esittelemät vaiheet tuovat tekijänoikeudellisen relevanssin *Aggarwalin* datatieteelliseen ja yleistasoiseen hahmotteluun. Pelkästään datatieteelliseen näkökulmaan keskittyminen saattaisi johtaa tekijänoikeudellisen

⁴⁵ ”Tieto” ja ”Informaatio”, Tieteen termipankki.

⁴⁶ Aggarwal 2015, s. 3-4.

⁴⁷ Rosati 2019, s. 8.

perspektiivin kadottamiseen tiedonlouhintaprosessin pääpiirteiden analysoinnista. Näin ollen seuraavaksi käsitellään tekstin- ja tiedonlouhintaprosessien keskeisiä vaiheita limittäin yhtäältä datatieteilijän ja toisaalta oikeustieteen asiantuntijan näkövinkkelistä katsoen. Käsittelyä täydennetään soveltuvien osin myös muun relevantin aineiston avulla.

Rosatin mainitsema pääsy sisältöön tulkitaan myös tässä tutkielmassa ensimmäiseksi vaiheeksi tekijänoikeudellisesti relevanttia tekstin- ja tiedonlouhintaprosessia. Vaihe saattaa ensi lukemalta vaikuttaa itsestään selvältä tai jopa merkityksettömältä, mutta on tärkeää huomioida, että kaikkiin sisältöihin ei ole vapaata pääsyä. Mikäli sisältöön ei pääse käsiksi, sitä on mahdotonta hyödyntää tekstin- ja tiedonlouhinnan lähdemateriaalina. Rosati on jakanut erilaiset sisällöt karkeasti kahtia sen mukaan, onko niihin vapaa pääsy vai ei. Jälkimmäiseen ryhmään kuuluviin sisältöihin pääseminen saattaa vaatia esimerkiksi lisenssin hankkimista.⁴⁸

Edellä kuvattua erityyppisten sisältöjen kahtiajakoa voidaan havainnollistaa esimerkiksi siten, että joitakin kirjallisuuden teoksia voi lukea maksutta ja vapaasti, kun taas toisiin teoksiin käsiksi pääseminen perustuu luvan saamiseen ja mahdolliseen maksullisuuteen lisenssijärjestelmien kautta. On kuitenkin syytä huomauttaa, että kummalla tahansa keinolla hankittu pääsy sisältöön ei kategorisesti tarkoita sitä, että sisältö olisi sen jälkeen vapaasti louhittavissa. Näin on siitä huolimatta, että ”oikeuden lukea tulisi tarkoittaa oikeutta louhia” (right to read should be the right to mine) -tyylinen lähestymistapa onkin saanut suhteellisen laajasti tuulta siipiensä alle.⁴⁹

Pääsy sisältöön on yhteydessä Aggarwalin mainitsemaan datan keräämiseen, sillä mikäli sisältöön ei ole pääsyä, sitä ei voi myöskään kerätä louhintaa varten. Lieneekin koherenttia käsitellä datan keräämistä vasta nyt, kun oletuksena on, että kerättyihin sisältöihin on päästy käsiksi tavalla tai toisella. Dataa voidaan kerätä monin eri tavoin. Toiston ja liiallisen luetteloinnin välttämiseksi mainittakoon esimerkinomaisesti, että dataa voidaan kerätä sekä sitä varten erityisesti kehitetyn laitteiston, kuten sensorien, avulla että manuaalisena työnä, jolloin keräämisessä voidaan hyödyntää vaikkapa kyselytutkimusta.⁵⁰ Edellä mainitut kirjallisuuden teokset voivat toimia myös louhinnan raaka-aineena. Tällöin louhinnan

⁴⁸ Rosati 2019, s. 8.

⁴⁹ Geiger – Frosio – Bulayenko 2018, s. 27.

⁵⁰ Aggarwal 2015, s. 3.

tarkoituksena voi olla esimerkiksi selvittää tiettyjen sanojen käytön yleistymistä tai harvenemista vuosikymmenten tai -satojen kuluessa.

Nähdäkseni Rosatin mainitsema sisällön eriyttämistä ja/tai kopiointia sisältävä tekstin- ja tiedonlouhintaprosessin vaihe nivoutuu yhteen sekä sisältöön pääsyn että datan keräämisen kanssa. Näkökulmaa voitaneen perustella edelleen kirjallisuuden alaa esimerkkinä käyttäen. Mikäli louhittavaksi suunniteltu tietomuoto on tekstiä, ja tavoitteena on vaikkapa analysoida kielen kehittymistä tietyllä aikajaksolla, louhinnan raaka-aineena käytettäneen kirjallisuuden teoksia. Manuaalisena työnä projektiin kului valtavasti aikaa, mutta oikeanlaista tiedonlouhinta-algoritmia hyödyntäen se voidaan toteuttaa varsin nopeasti. Algoritmin hyödyntämiseksi teokset tulee kuitenkin muun muassa muuntaa koneluettavaan muotoon, mikä edellyttää teosten tekstin kopiointia ja eriyttämistä samaan tietovarastoon (data warehouse), jossa niitä voidaan käsitellä. Tietokantojen osalta sisällön eriyttäminen ja/tai kopiointi saattaa puolestaan tarkoittaa sitä, että valitaan laajasta tietokannasta kussakin tilanteessa kiinnostavimmat osat, jotka siirretään niille erikseen luotavaan pienempään tietokantaan.⁵¹ Joissain tapauksissa voi myös olla niin, että analysoitavaksi haluttu data sijaitsee eri tietokannoissa, jolloin kyseiset tiedot täytyy luonnollisesti yhdistää toisiinsa uudessa tietokannassa.⁵²

Sisällön eriyttäminen ja/tai kopiointi lienee ainakin tekijänoikeudelliselta kantilta katsottuna tekstin- ja tiedonlouhintaprosessin tärkein vaihe, sillä sen tullessa ajankohtaiseksi joudutaan usein tilanteeseen, jossa näiden toimenpiteiden suorittamiseen vaaditaan erillinen lupa tekijänoikeuden haltijalta tai tietokannan luojalta.⁵³ Syy luvanvaraisuuteen on se, että toimenpiteet merkitsevät juridisesti tulkittuna kappaleen valmistamista.

Kappaleen valmistaminen voi olla tekstin- ja tiedonlouhintaprosessissa luonteeltaan joko tilapäistä tai pysyvää. Tilapäinen kappaleen valmistaminen voi perustua jo yksinään haun suorittamiseen tietokannassa, kun taas pysyvä kappaleen valmistaminen voi liittyä esimerkiksi toimenpiteeseen, jossa louhittavaksi suunnitellut sisällöt siirretään niiden alkuperäisistä sijainneista omaan uuteen tietokantaansa.⁵⁴ Kappaleen valmistamista koskevaa oikeutta on säännelty sekä tekijänoikeuslainsäädännössä että tietokantojen *sui*

⁵¹ Alpaydin 2016, s. 154.

⁵² Alpaydin 2016, s. 154.

⁵³ Rosati 2019, s. 10.

⁵⁴ Flynn – Geiger – Quintais ym. 2020, s. 394.

generis -suojaa koskeissa säännöksissä. Tähän sääntelyyn tullaan pureutumaan tarkemmin seuraavassa pääluvussa.

Kerätty data on harvoin louhintakelpoista sellaisenaan. Tällainen raakadata on usein valmistelematonta, vaillinaista, vioittunutta tai vääristynyttä.⁵⁵ Tämän vuoksi dataa täytyy esikäsitellä ennen louhinnan toteuttamista. Esikäsitelyllä tarkoitetaan tässä yhteydessä tiedon relevanttien ominaisuuksien eriyttämistä (feature extraction), tiedon puhdistamista (data cleaning) sekä tiedon ominaisuuksien valintaa ja niiden muokkaamista (feature selection and transformation). Puhuttaessa tiedon relevanttien ominaisuuksien eriyttämisestä tarkoitetaan tiedonlouhinnan tarkoituksen kannalta hyödyllisten tiedon ominaisuuksien erottamista sellaisista tiedon ominaisuuksista, joita ei nähdä yhtä merkityksellisiksi. Esimerkiksi luottokorttihuijauksia tutkittaessa veloituksen määrä, veloitusten säännöllisyys ja sijainti ovat hyödyllisiä toimintoja, kun taas jotkin muut toiminnot saattavat olla käytännössä hyödyttömiä.⁵⁶

Relevanttien toimintojen eriyttämistä seuraa tiedon puhdistaminen. Tiedon puhdistamiselle voi olla erilaisia, tietomuodoista riippuvaisia syitä. Tutkielman viitekehyksessä syistä voidaan nostaa erityisesti esille tilanteet, joissa kerätyn tiedon havaitaan olevan väärää tai epätarkkaa.⁵⁷ Esimerkiksi edellä mainitun The Next Rembrandt -projektin yhteydessä Rembrandtin tuotantoa analysoitiin pikseli pikseliltä, minkä vuoksi tiedon valmisteluvaiheista etenkin puhdistamisen voidaan olettaa olleen merkityksellisessä asemassa. Väärän tai epätarkan tiedon puhdistamista on syytä korostaa myös siitä syystä, että mikäli tällaista tietoa ei puhdisteta, tiedon myöhempi analysointi perustuu yhtä lailla näiden virheellisyyksien varaan. Ihmisten elämään vaikuttavan tekoälysovelluksen kehittäminen tällaiselta pohjalta voisi osoittautua erittäin haitalliseksi.

Tiedon ominaisuuksien valinta ja niiden muokkaaminen muodostavat tiedon esikäsitelyvaiheen viimeisen osan, vaikka ne voidaan useissa tapauksissa linkittää itse louhinnassa käytettäviin algoritmeihin ja siten erillisiksi valmisteluprosessista.⁵⁸ Tiedon ominaisuuksien valinnassa on jo sanamuodollisestikin tulkittuna kyse käytännössä tiedon relevanttien ominaisuuksien eriyttämisen pidemmälle viedystä sovellutuksesta. Tällöin ominaisuuksia voidaan edelleen karsia esimerkiksi siihen perustuen, mitkä niistä tuottavat

⁵⁵ Larose, D. – Larose, C. 2014, s. 17.

⁵⁶ Aggarwal 2015, s. 3.

⁵⁷ Geetha ym. 2019, s. 776.

⁵⁸ Aggarwal 2015, s. 6.

eniten hyödyllistä informaatiota tiettyä tiedonlouhinta-algoritmia käytettäessä.⁵⁹ Toimintojen muokkaamisessa on puolestaan kyse esimerkiksi tietyn toiminnon (esim. ikä) jakamisesta isompaan väliin.⁶⁰

Aggarwalin jaottelun viimeisessä vaiheessa tekstin- ja tiedonlouhintaprosessissa edetään tiedon analysointiin. Analysointivaiheessa tieto syötetään ohjelmistolle, jonka toiminta perustuu valittuun tiedonlouhinta-algoritmiin eli -menetelmään. Algoritmin tarkoituksena on tuottaa tiedosta hahmoja tai malleja, joita tulkitsemalla mahdollistetaan uuden tiedon luominen ja oivallusten tekeminen.⁶¹ Jotta tarkoitus voidaan täyttää mahdollisimman hyvin, on valittava tiedonlouhinnan soveltamiskohteeseen luontevimmin sopiva algoritmi. Rosatin luokittelun kolmas vaihe perustuu samoihin kulmakiviin kuin Aggarwalilla. Rosati kuitenkin kutsuu tiedon analysointivaihetta varsinaiseksi tiedon louhinnaksi.⁶² Tämä vaihe on puolestaan jaettu edelleen kahteen vaiheeseen, joista ensimmäinen koskee datan esikäsittelyä ja toinen strukturoidun datan eriyttämistä.⁶³

Edellä kuvatut semanttiset eroavaisuudet tekstin- ja tiedonlouhintaprosessin kuvailussa eivät ole tämän tutkielman kannalta kovinkaan ratkaisevassa osassa. Esiteltäköön aihekokonaisuutta kuitenkin lyhyesti tässä yhteydessä. Tekstin- ja tiedonlouhinta -termiä voidaan pitää joko aina datan keräämisestä uusien oivallusten esittämiseen saakka ulottuvan prosessin yläkäsitteenä tai ainoastaan yhtenä välivaiheena tässä prosessissa, jota kutsutaan tällöin kokonaisuudessaan tietämyksen lisäämiseksi (knowledge discovery from data).⁶⁴ On kuitenkin niin, että eri mallien vaiheet vastaavat hyvin pitkälti toisiaan, vaikka jälkimmäisessä mallissa datan analysointia kutsutaankin datan louhinnaksi ja ensimmäisessä louhintaa koko prosessia kuvaavaksi termiksi.

2.3 Tekstin- ja tiedonlouhinnan suhde koneoppimiseen ja tekoälyyn

Edellisessä alaluvussa kuvattiin tekstin- ja tiedonlouhintaprosessin piirteitä sekä datatieteen ylätasolla että tekijänoikeudellisesti relevanteilta osin. Ennen tarkempaa syventymistä tekstin- ja tiedonlouhinnan sääntelykehikkoon on tutkielman ymmärrettävyyden kannalta olennaista perehtyä lyhyesti tekstin- ja tiedonlouhintaan myös koneoppimisen ja tekoälyn

⁵⁹ Bramer 2016, s. 149.

⁶⁰ Aggarwal 2015, s. 5.

⁶¹ Bramer 2016, s. 2-3.

⁶² Rosati 2019, s. 12-13.

⁶³ Rosati 2019, s. 12-13.

⁶⁴ Han – Kamber – Pei 2011, s. 6.

konteksteissa. Koneoppiminen ja tekoäly mahdollistavat useita jo tutkielman johdantoluvussakin mainituista yhteiskunnallisesti merkittävistä keksinnöistä, minkä vuoksi on johdonmukaista pyrkiä sijoittamaan tekstin- ja tiedonlouhinta osaksi niiden muodostamaa kokonaisuutta.

Yksi koneoppimisen varhaisista määritelmistä esitettiin jo vuonna 1959. Määritelmän mukaan koneoppiminen on tutkimusala, joka antaa tietokoneelle kyvykkyyden oppia siitä huolimatta, että tietokonetta ei ole eksplisiittisesti ohjelmoitu.⁶⁵ Koneoppimisella tarkoitetaan hieman yksityiskohtaisemmin ilmaistuna koneen ohjelmointia siten, että se kykenee täyttämään sille annetun tehtävän kriteerit mahdollisimman optimaalisesti käyttäen apunaan kerättyä dataa tai aiempaa kokemusta.⁶⁶

Käytettäessä koneoppivaa algoritmia koneelle ei syötetä ihmisen laatimia vaihteellaisia ohjeita, joissa määriteltäisiin koneen noudattama toimintapa erilaisissa skenaarioissa. Sen sijaan koneelle annetaan tavoite tietystä lopputuloksesta, ja kone saa itse ratkaista parhaan tavan saavuttaa tämä tavoite.⁶⁷ Tässä mielessä koneoppivien algoritmien tapa oppia muistuttaa ihmisten ja muiden elävien olentojen yleistä oppimismetodia. Koneoppivia algoritmeja on monenlaisia, ja tietyn algoritmin käytön valitseminen perustuu sen soveltamiskohteeseen.

Koneoppimista ja tiedonlouhintaa saatetaan joissakin yhteyksissä pitää synonyymeina keskenään. Tiedonlouhinnan on myös esitetty olevan yksi koneoppimisen muodoista.⁶⁸ *Gollapudi* on kuitenkin katsonut, että vaikka koneoppimisen ja tiedonlouhinnan tavoite uuden tiedon tuottamisesta kerätyn datan pohjalta onkin sama, niillä on myös yksi eroavaisuus keskenään: siinä missä koneoppiminen keskittyy käyttämään tunnettua tietämystä ja kokemusta datan analysoinnissa, tiedonlouhinnassa päähuomion saa uuden tietämyksen löytäminen kerätystä datasta.⁶⁹ Tässä tutkielmassa tiedonlouhinnan ja koneoppimisen rajanveto tehdään juuri tuon nyanssieron varaan. Tällöin koneoppiminen voidaan hahmottaa ikään kuin ylimääräiseksi analyysikerrokseksi, joka on rakennettu tiedonlouhinnan avulla kootun tietämyksen jatkeeksi.⁷⁰

⁶⁵ Artificial Intelligence, Robotics, Privacy and Data Protection 2016, s. 19.

⁶⁶ Alpaydin 2014, s. 3.

⁶⁷ Fry 2018, s. 24.

⁶⁸ Alpaydin 2016, s. 14.

⁶⁹ Gollapudi 2016, s. 30.

⁷⁰ Flynn – Geiger – Quintais ym. 2020, s. 394.

Edellisessä kappaleessa todetun tarkennukseksi sekä tiedonlouhinnan ja koneoppimisen välisen suhteen selvennykseksi voidaan lausua, että tässä tutkielmassa koneoppimisella tarkoitetaan hieman yksinkertaistaen koneoppivia algoritmeja, jotka saavat harjoitusdatansa tekstin- ja tiedonlouhinnan avulla kootusta tietämyksestä. Tällöin tekstin- ja tiedonlouhintaa käytetään Gollapudinkin mainitsemalla tavalla uuden tietämyksen löytämiseen kerätystä datasta, kun taas koneoppiminen perustuu tämän hankitun tietämyksen syöttämiseen algoritmeille, jotka pyrkivät harjoitusdatan avulla oppimaan suorittamaan uusia tehtäviä, joiden suorittamiseen niitä ei ole alun perin spesifisti ohjelmoitu.

Tekoälystä ja sen kasvavasta merkityksestä nyky-yhteiskunnalle on käyty alati kiihtyvää keskustelua viimeisen vuosikymmenen aikana. Ei liene kärkeä todeta, että osa tästä keskustelusta on ajautunut ajoittain väärille urille ja sivuuttanut täysin sen, mitä tekoälyllä käytännössä oikeastaan tarkoitetaan. Tämän tutkielman intressien kannalta ei ole perusteltua ajautua syvälle diskurssiin älykkyyden tarkasta määritelmästä. Todettakoon kuitenkin, että niin kutsuttujen klassisten tekoälysovellusten pyrkimyksenä oli saada kone kopioimaan ihmisälykkyyttä.⁷¹ Tällöin lähtökohtana oli katsoa ihmisaivojen toimintaa ulkopuolelta ja pyrkiä tehtyjen havaintojen pohjalta saada tekoälyjärjestelmä toistamaan aivojen suoritus.⁷² Monilla lajeilla älykkyyden peruspiirteisiin kuuluu kuitenkin myös erilaisten tilanteiden hahmottaminen ja niiden vertaaminen aiemmin opittuihin kokemuksiin. Tällaiset piirteet auttavat meitä havaitsemaan, mikä toimii ja mikä ei sekä toimimaan näiden havaintojen ja kokemusten pohjalta uusissa tilanteissa niin hyvin kuin mahdollista. Moderneja tekoälyjärjestelmiä kehitetäänkin juuri tällaisen kokonaisvaltaisemman aivojen toiminnan tutkimisen kautta.⁷³

Koneoppimista voidaan pitää tekoälyn alalajina, jolla on ollut huomattavan suuri merkitys tekoälykehityksessä.⁷⁴ Tämä perustuu yhtäältä koneoppivien järjestelmien edellä esitettyyn kyvykkyyteen oppia ilman, että niitä on nimenomaisesti ohjelmoitu ja toisaalta siihen, että järjestelmät kehittyvät automaattisesti kokemuksen myötä.⁷⁵ Nähdäkseni tässä kokonaisuudessa modernit tekoälyjärjestelmät ja koneoppiminen nivoutuvat yhteen. Modernien tekoälyjärjestelmien edellä kuvatun kehittämismetodin voitaneen nimittäin katsoa ainakin osittain perustuvan koneoppimiselle tyypilliseen pyrkimykseen oppia sille

⁷¹ Warwick 2011, s. 31.

⁷² Warwick 2011, s. 88.

⁷³ Warwick 2011, s. 89.

⁷⁴ Kamarinou ym. 2016, s. 3.

⁷⁵ Mitchell 1997, s. XV.

syötetyistä aiemmista kokemuksista eli harjoitusdatasta suoriutuakseen jatkossa optimaalisemmin tietyssä tehtävässä. Tältä pohjalta voidaankin argumentoida suuren osan nykyajan tekoälystä tarkoittavan käytännössä erilaisten koneoppivien algoritmien soveltamista.

Mihin tekstin- ja tiedonlouhinta sitten sijoittuu suhteessa tekoölyyn? On esitetty, että tiedonlouhinta on kehittynyt tekoölyn alalajina suunnilleen 1990-luvun alusta alkaen.⁷⁶ Tässä mielessä tekoälyä voidaan pitää koneoppimisen ohella myös tekstin- ja tiedonlouhinnan yläkategoriana. Näin ollen tekstin- ja tiedonlouhinnan suhde tekoölyyn voidaan nähdäkseni määritellä edellä käydyn koneoppimisen ja tiedonlouhinnan rajapinta-analyysin kautta. Tällöin lähtökohtana on, että tiedonlouhinta mahdollistaa tekoölylle välttämättömän tietämyksen hankkimisen suurista tietokannoista.⁷⁷ Tällainen tietämyksen hankkiminen on kuitenkin myös koneoppimisen taustalla, sillä koneoppivat algoritmit voivat käyttää tekstin- ja tiedonlouhintaprosessissa puhdistettuja ja kohdennettuja datajoukkoja sekä niistä johdettuja trendejä ja kaavoja harjoitusdatana, jonka avulla ne voivat oppia ennustamaan tulevaisuuden tapahtumia tai tehdä muutoin päätöksiä epävarmuuden vallitessa.⁷⁸ Näin ollen tekoäly hahmotetaan tässä tutkielmassa eräänlaiseksi kattotermiksi, jonka isona osa-alueena on koneoppiminen. Tekstin- ja tiedonlouhinta on puolestaan onnistuneiden koneoppivien algoritmien kehittämisen ja siltä osin myös tekoälykehityksen välttämätön edellytys.

⁷⁶ Ertel 2017, s. 10.

⁷⁷ Ertel 2017, s. 10.

⁷⁸ Murphy 2012, s. 2.

3. Tekstin- ja tiedonlouhinnan sääntely EU:ssa

3.1 DSM-direktiiviä edeltänyt oikeustila

Tutkielman ensimmäinen tutkimuskysymys pureutuu siihen, millä tavoin tekstin- ja tiedonlouhinnan immateriaalioikeudellinen sääntely muuttuu EU:ssa DSM-direktiivin säätämisen myötä. Näin ollen ennen itse DSM-direktiivin syvällisempää analysointia on syytä tarkastella tekstin- ja tiedonlouhinnan relevanttia EU-oikeudellista sääntelykehikkoa aikana ennen DSM-direktiiviä. Mainittuna aikakautena tekstin- ja tiedonlouhinnan sääntely nojasi pääasiassa tietoyhteiskuntadirektiivin, tietokantadirektiivin ja tietokoneohjelmadirektiivin varaan. Seuraavissa alaluvuissa esitellään näiden direktiivien tekstin- ja tiedonlouhinnan kannalta olennaiset säännökset ja asetetaan ne louhintaprosesseille tyypillisten vaiheiden kontekstiin. Samassa yhteydessä esitetään myös alustavia tulkintoja mainittujen kolmen direktiivin soveltuvuudesta tekstin- ja tiedonlouhinnan sääntelyinstrumenteiksi.

Peilattaessa tietoyhteiskuntadirektiiviä, tietokantadirektiiviä ja tietokoneohjelmadirektiiviä tekstin- ja tiedonlouhintaan tutkielmassa lähdetään siitä oletuksesta, että louhittavaksi suunniteltuun sisältöön on päästy käsiksi laillisin keinoin. Pääsy tekijänoikeudella tai muulla immateriaalioikeudella suojattuun sisältöön ei nähdäkseni ole tekstin- ja tiedonlouhinnan kontekstissa juridisesti kompleksisempi aihekokonaisuus kuin muutkaan teosten käyttötavat. Kyse on käytännössä samasta tilanteesta kuin esimerkiksi haluttaessa lukea tiettyä julkaistua sisältöä: ennen lukemista on saatava pääsy kyseessä olevaan sisältöön. Tässä yhteydessä voidaan muodostaa jako yhtäältä vapaasti saatavilla olevaan sisältöön ja toisaalta sisältöön, johon pääsyä on rajoitettu. Pääsyä voidaan rajoittaa esimerkiksi lisenssin hankkimista edellyttämällä.⁷⁹

3.1.1 Tietoyhteiskuntadirektiivi

Tietoyhteiskuntadirektiivi on ollut pian voimassa jo 20 vuotta, ja sitä on kuvailtu muun muassa EU:n historian lobatuimmaksi direktiiviksi.⁸⁰ Riippumatta kuvauksen paikkansapitävyydestä tietoyhteiskuntadirektiivin voidaan ainakin aikalaislähteiden perusteella katsoa olleen ensimmäinen EU-direktiivi, joka harmonisoi tekijänoikeutta ja lähioikeuksia koskevia kysymyksiä laajemmasta ja horisontaalisesta näkökulmasta.⁸¹

⁷⁹ Aihekokonaisuutta on käsitelty tarkemmin luvussa 2.

⁸⁰ Shapiro 2019, s. 44.

⁸¹ Geiger – Schönherr 2014, s. 397.

Harmonisointi pyrittiin toteuttamaan korvaamalla laajoja osia eri jäsenvaltioiden kansallisista tekijänoikeussäädöksistä tietoyhteiskuntadirektiivin yleiseurooppalaisiksi tarkoitetuilla säännöksillä.⁸² Kuten myöhemmin käsillä olevassa pääluvussa tullaan toteamaan, harmonisointitavoitteessa onnistuttiin tietoyhteiskuntadirektiivin osalta ainoastaan paikoittain.

Tietoyhteiskuntadirektiivin tekstin- ja tiedonlouhinnan kannalta merkityksellinen aines sisältyy 2 artiklassa säänneltyyn kappaleen valmistamista koskevaan oikeuteen sekä joihinkin 5 artiklassa luetelluista poikkeuksista ja rajoituksista kappaleen valmistamista koskevaan oikeuteen. Tämä johtuu siitä jo edellä tutkielmassa havainnollistetusta lähtökohdasta, että useissa tekstin- ja tiedonlouhintaprosesseissa louhittavaksi suunniteltua sisältöä täytyy eriyttää ja/tai kopioida. Tällaisten toimenpiteiden voidaan melko kategorisesti katsoa lukeutuvan tekijänoikeuden haltijalle kuuluvan kappaleen valmistamista koskevan yksinoikeuden piiriin. Lieneekin perusteltua todeta, että tietoyhteiskuntadirektiivi oli DSM-direktiiviä edeltäneen ajan tekstin- ja tiedonlouhintaa koskeneen immateriaalioikeussääntelyn kulmakivi.

Tietoyhteiskuntadirektiivin 2 artiklan mukaan jäsenvaltioiden on säädettävä, että yksinoikeus sallia tai kieltää suoraan tai välillisesti, tilapäisesti tai pysyvästi, millä keinolla ja missä muodossa tahansa kokonaan tai osittain tapahtuva kappaleen valmistaminen on muun muassa tekijöillä teostensa osalta. Esimerkiksi Suomen tekijänoikeuslain osalta mainittu artikla heijastuu erityisesti 2 §:ään, jonka 1 momentissa kerrotaan, että tekijänoikeus tuottaa, jäljempänä säädetyin rajoituksin, yksinomaisen oikeuden määrätä teoksesta muun muassa valmistamalla siitä kappaleita. Saman pykälän 2 momentissa puolestaan tarkennetaan kappaleen valmistamisen juridista merkityssisältöä seuraavasti: ”Kappaleen valmistamisena pidetään sen valmistamista kokonaan tai osittain, suoraan tai välillisesti, tilapäisesti tai pysyvästi sekä millä keinolla ja missä muodossa tahansa. Kappaleen valmistamisena pidetään myös teoksen siirtämistä laitteeseen, jolla se voidaan toisintaa.”

Tietoyhteiskuntadirektiiviin sisältyy myös rajoituksia ja poikkeuksia, joista joitakin voidaan pitää tekstin- ja tiedonlouhinnan näkökulmasta relevantimpina kuin toisia. Direktiivin 5 artiklan 1 kohdan mukaan 2 artiklassa säädetty kappaleen valmistamista koskeva oikeus ei

⁸² Guze 2019, s. 71.

koske sellaista 2 artiklassa tarkoitettua tilapäistä kappaleen valmistamista, joka on väliaikaista tai satunnaista sekä erottamaton ja välttämätön osa teknistä prosessia ja jonka ainoa tarkoitus on mahdollistaa: a) välittäjän toimesta tapahtuva teoksen tai muun aineiston siirto verkossa kolmansien osapuolien välillä, tai b) sen laillinen käyttö, ja jolla ei ole itsenäistä taloudellista merkitystä.

Tietoyhteiskuntadirektiivin 5 artiklan 2 kohdassa on puolestaan listattu tapauksia, joissa jäsenvaltiot voivat säätää poikkeuksista tai rajoituksista 2 artiklassa säädettyyn kappaleen valmistamista koskevaan oikeuteen. Tutkielman tarkoituksen kannalta ei ole tarpeellista eritellä mainitussa kohdassa listattuja tapauksia yksityiskohtaisesti. Riittääkin todeta, että mikään niistä ei mahdollista välillisesti tai välittömästi kaupallisessa tarkoituksessa toteutettavaa tekstin- ja tiedonlouhintaa. Sama pätee myös tietoyhteiskuntadirektiivin 5 artiklan 3 kohtaan, joka sekin sisältää kuvattun tyyppisen tapauslistauksen, joskin kohdistuen kappaleen valmistamista koskevan oikeuden lisäksi myös direktiivin 3 artiklassa säädettyyn oikeuteen välittää yleisölle teoksia ja oikeuteen saattaa muu aineisto yleisön saataviin.

Tietoyhteiskuntadirektiivin 5 artiklan 2 ja 3 kohdat toimivat nähdäkseni kuitenkin esimerkkeinä direktiivin harmonisointitavoitteen epäonnistumisesta. Kohdissa käytetty sanamuoto ”jäsenvaltiot voivat” kertoo, että jäsenvaltioiden ei ole ollut pakko implementoida listattuja tapauksia kappaleen valmistamista koskevan oikeuden poikkeuksiksi tai rajoituksiksi. Tämän seurauksena eri jäsenvaltiot ovat noudattaneet toisistaan eriäviäisiä sääntelyratkaisuja.⁸³ On selvää, että tekijänoikeussäätelyn koko EU-alueen kattava harmonisointi edellyttäisi kuvattua vahvempia toimenpiteitä.

Tietoyhteiskuntadirektiivin 5 artiklan 5 kohdan mukaan artiklan aiempien kohtien mukaisia poikkeuksia ja rajoituksia sovelletaan vain tietyissä erityistapauksissa, jotka eivät ole ristiriidassa teoksen tai muun aineiston tavanomaisen hyödyntämisen kanssa eivätkä kohtuuttomasti haittaa oikeudenhaltijan oikeutettuja etuja. Kohta perustuu niin kutsuttuun kolmivaihetestiin (three-step test), joka omaksuttiin osaksi tietoyhteiskuntadirektiiviä muun muassa Bernin sopimuksen⁸⁴, TRIPS-sopimuksen (Agreement on Trade-Related Aspects of

⁸³ Bottis ym. 2019, s. 191.

⁸⁴ Bernin yleissopimus kirjallisten ja taiteellisten teosten suojaamisesta.

Intellectual Property Rights)⁸⁵, WIPO:n tekijänoikeussopimuksen ja WIPO:n esitys- ja äänitesopimuksen kautta.⁸⁶

Kolmivaihetesti perustuu nimensä mukaisesti kolmeen eri edellytykseen myös tietoyhteiskuntadirektiiviin implementoidussa muodossaan. Edellytykset onkin helppo eritellä direktiivin 5 artiklan 5 kohdan tekstistä: 1. poikkeukset ja rajoitukset soveltuvat vain tiettyihin erityistapauksiin; 2. nämä poikkeukset tai rajoitukset eivät saa olla ristiriidassa teoksen tai muun aineiston tavanomaisen hyödyntämisen kanssa; 3. eivätkä ne saa myöskään kohtuuttomasti haitata oikeudenhaltijan oikeutettuja etuja.

Pihlajarinne on tulkinnut, että kolmivaihetestin ensimmäinen edellytys korostaa tekijänoikeuden rajoitusten ja poikkeusten suppeaa tulkintaa, kun taas toisen ja kolmannen edellytyksen tavoitteena on ennemmin eri intressien tasapainottaminen.⁸⁷ Nähdäkseni on kuitenkin niin, että kaikki edellytykset painottavat sanamuodoissaan selkeämmin oikeudenhaltijoiden kuin teoksia käyttävän yleisön etunäkökohtia. Tämä on mielenkiintoista, sillä tietoyhteiskuntadirektiivin johdanto-osan 31 perustelukappaleessa ilmaistaan tavoitteeksi muun muassa eri oikeudenhaltijaryhmien ja suojatun aineiston käyttäjien välisten oikeuksien ja etujen oikeudenmukaisen tasa-arvon turvaaminen.

On syytä huomauttaa, että kolmivaihetestin omaksuminen osaksi tietoyhteiskuntadirektiiviä ei merkinnyt pyrkimystä harmonisoida tekijänoikeuspoikkeusten alaa EU-maiden kansallisissa laeissa. Sen sijaan tarkoituksena oli luoda eräänlainen johtava periaate, jonka käyttö nähtiin jäsenvaltiotasolla kansallisen lain salliessa mahdolliseksi tapauksissa, joiden ratkaisun tueksi ei ollut löydettävissä spesifiä poikkeussäännöstä.⁸⁸

Tekstin- ja tiedonlouhintaprosessit saattavat aiempaan tutkielmassa kuvatuin tavoin loukata teoksen tekijänoikeuden haltijan yksinoikeutta kappaleen valmistamiseen silloin, kun louhittavaksi suunniteltu sisältö ylittää teoskynnyksen ja on siten tekijänoikeussuojan alainen.⁸⁹ Tällöin on huomioitava, että jopa hyvin lyhyen otteen kopioiminen tekijänoikeudella suojatusta teoksesta voidaan luokitella oikeudenhaltijan yksinoikeuden loukkaukseksi. Tämä käy ilmi *Infopaq*-tapauksesta, jonka tuomion perusteluissa katsottiin, että yhdentoista sanan mittainen ote suojatusta teoksesta voi kopioituna loukata

⁸⁵ Sopimus teollis- ja tekijänoikeuksien kauppaan liittyvistä näkökohdista.

⁸⁶ Mazziotti 2008, s. 80.

⁸⁷ Pihlajarinne 2004, luku V: Immateriaalioikeudet.

⁸⁸ Gervais 2017, s. 73.

⁸⁹ Geiger ym. 2018, s. 6.

oikeudenhaltijan yksinoikeutta kappaleen valmistamiseen.⁹⁰ Näkökulmaa tukee myös tietoyhteiskuntadirektiivin johdanto-osan 21 perustelukappale, jonka mukaan kappaleen valmistamista koskevaa oikeutta tulisi tulkita laajasti. Yksinoikeutta loukkaavaa kopiointia voi tapahtua esimerkiksi standardisoitaessa sisältöä koneluettavaan muotoon ja tiettyjä louhintatekniikoita käytettäessä myös vaiheessa, jossa esikäsitelty sisältö ladataan sille varatulle alustalle.⁹¹

Tietoyhteiskuntadirektiivin kappaleen valmistamista koskevaa oikeutta rajoittavista tai siihen poikkeuksia asettavista säännöksistä ainoastaan 5 artiklan 1 kohta on laadittu jäsenvaltioille pakollisesti implementoitavaksi. Tekstin- ja tiedonlouhinnan kannalta säännöksestä seuraa, että vaikka louhintaprosessiin sisältyisikin vaihe, jossa louhittavaksi suunniteltua sisältöä jouduttaisiin eriyttämään ja/tai kopioimaan, nämä toimet voisivat ainakin teoriassa olla ilman oikeudenhaltijan hyväksyntääkin luvallisia tiettyjen edellytysten täytyessä. Myös *Rosati* on katsonut, että poikkeussäännös mahdollistaisi sen asettamissa puitteissa tapahtuvan tilapäisen kopioinnin tekstin- ja tiedonlouhintaprosessin nimissä.⁹²

Tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan poikkeussäännöstä on täsmennetty oikeuskäytännössä jonkin verran. *Premier League* -tapauksen tuomion perusteluissa määritellään säännöksen tavoite seuraavalla tavalla: ”... poikkeuksen on tavoitteensa mukaisesti siten mahdollistettava ja taattava uusien teknologioiden kehittyminen ja toimivuus sekä säilytettävä yhtäältä oikeudenhaltijoiden oikeuksien ja etujen sekä toisaalta suojatun aineiston käyttäjien, jotka haluavat käyttää näitä uusia teknologioita, välisten oikeuksien ja etujen oikeudenmukainen tasapaino”.⁹³ Säännöksen soveltumisvaatimuksia on täsmennetty portaittain sekä *Infopaq I* ja *II* -tapauksissa että edellä mainitun *Premier League* -tapauksen tuomion perusteluissa.

Infopaq I -tapauksen tuomion perusteluiden mukaan kappaleen valmistamista voidaan pitää väliaikaisena vain, ”jos sen olemassaoloaika on rajoitettu siihen, mikä on kyseisen teknisen prosessin asianmukaisen toiminnan kannalta tarpeen, edellyttäen, että kyseisen teknisen prosessin on oltava sillä tavoin automaattinen, että valmistettu kappale poistetaan automaattisesti ilman ihmisten toimintaa heti, kun se on täyttänyt tehtävänsä, jonka

⁹⁰ C-5/08 (*Infopaq*), kohdat 54-55.

⁹¹ Geiger ym. 2018, s. 7.

⁹² *Rosati* 2019, s. 10.

⁹³ C-403/08 ja C-429/08 (*Premier League*), kohta 164.

tarkoituksena on mahdollistaa tällaisen prosessin saattaminen loppuun”.⁹⁴ Kappaleen valmistamisen satunnaisuutta on puolestaan tulkittava yhteydessä teoksen hyödyntämisen kokonaisuuteen; satunnainen kappaleen valmistaminen on kyseessä silloin, kun kopio on satunnainen suhteessa teoksen pääasialliseen hyödyntämistapaan nähden.⁹⁵

Kappaleen valmistamisen tuloksena syntyvän kopion luonne erottamattomana ja välttämättömänä osana teknistä prosessia tarkoittaa yksinkertaistettuna sitä, että kopion täytyy mahdollistaa teoksen toisenlainen käyttö, joka toteutetaan teknologisen prosessin keinoin.⁹⁶ Euroopan unionin tuomioistuin on selventänyt kriteeriä Infopaq II -tapauksen tuomion perusteluissa, joissa todetaan seuraavasti: ”Käsite ’erottamaton ja välttämätön osa teknistä prosessia’ edellyttää sitä, että tilapäinen kappaleen valmistaminen tapahtuu tyystin teknistä prosessia täytäntöön pantaessa ja että sitä ei siis suoriteta kokonaan tai osittain tällaisen prosessin ulkopuolella. Kyseinen käsite edellyttää myös, että tilapäinen kappaleen valmistaminen on välttämätöntä siinä mielessä, että asianomainen tekninen prosessi ei voisi sujua asianmukaisesti ja tehokkaasti ilman kyseistä kappaleen valmistamista. Kun lisäksi otetaan huomioon se, että direktiivin 2001/29 5 artiklan 1 kohdassa ei täsmennetä sitä, missä vaiheessa teknistä prosessia tilapäisen kappaleen valmistamisen on tapahduttava, on mahdollista, että tällainen kappaleen valmistaminen aloittaa tai päättää kyseisen prosessin. Kyseisessä säännöksessä mikään ei myöskään osoita, ettei ihminen saisi toiminnallaan osallistua tekniseen prosessiin, eikä etenäkään, ettei kyseistä prosessia saisi käynnistää manuaalisesti ensimmäisen tilapäisen kappaleen valmistamiseksi.”⁹⁷

Tilapäisen kappaleen valmistaminen tarkoituksessa mahdollistaa välittäjän toimesta tapahtuva teoksen tai muun aineiston siirto verkossa kolmansien osapuolien välillä täsmentyä kriteerinä tietoyhteiskuntadirektiivin johdanto-osan 33 perustelukappaleessa. Perustelukappaleen mukaan tilapäisen kappaleen valmistamisen tulisi mahdollistaa siirtojärjestelmien tehokkaan toiminnan mahdollistava tallennus edellyttäen, että välittäjä ei muuta tietoja eikä puutu toimialalla laajalti hyväksytyn ja käytetyn teknologian lailliseen käyttöön saadakseen tietoja talletetun tiedon käytöstä. Samassa perustelukappaleessa määritellään myös laillisen käytön kriteeri poikkeussäännöksen alaisuudessa: ”Käyttöä olisi pidettävä laillisena, jos se on oikeudenhaltijan sallimaa tai jos sitä ei ole lailla rajoitettu”.

⁹⁴ C-5/08 (Infopaq), kohta 64.

⁹⁵ Triaille 2008, s. 257.

⁹⁶ Triaille ym. 2014, s. 43.

⁹⁷ C-302/10 (Infopaq II), kohdat 30-32.

Nämä kaksi poikkeussäännöksen sisältämää käyttötarkoitusta ovat tietoyhteiskuntadirektiivin sanamuodon perusteella toisistaan erillisiä siinä mielessä, että jommankumman niistä täytyminen riittää säännöksen soveltumiseen, mikäli muut kappaleen valmistamiselta edellytetyt kriteerit täyttyvät.

Tilapäisellä kappaleen valmistamisella ei saa tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan poikkeussäännöksen mukaan olla itsenäistä taloudellista merkitystä. Infopaq II -tapauksen tuomion perusteluissa tunnustetaan, että tilapäisellä kappaleen valmistamisella pyritään helpottamaan teoksen käyttöä tai tehdä sen käytöstä tehokkaampaa, minkä seurauksena toiminnan voitot kasvavat tai tuotantokustannukset laskevat. Tästä huolimatta Euroopan unionin tuomioistuin totesi, että ”tilapäisen kappaleen valmistamisella ei saa olla itsenäistä taloudellista merkitystä siinä mielessä, että tällaisen kappaleen valmistamisesta saatu taloudellinen etu ei saa olla asianomaisten teoksen laillisesta käytöstä saadusta taloudellisesta edusta eroava tai erotettavissa oleva etu eikä se saa johtaa sellaiseen taloudelliseen lisäetuu, joka ylittää mainitusta suojatun teoksen käytöstä saatavan edun”. Edelleen saman tapauksen tuomion perusteluissa täsmennetään kriteeriä seuraavasti: ”Tilapäisestä kappaleen valmistamisesta saatava etu on erilainen ja erotettavissa oleva, jos tilapäisen kappaleen valmistaja voi saada voittoa hyödyntämällä taloudellisesti tilapäisesti valmistettuja kappaleita itsessään. Samoin on silloin, kun tilapäinen kappaleen valmistaminen johtaa valmistetun aineiston, sellaisena kuin se on olemassa asianomaista teknistä prosessia käynnistettäessä, muuttamiseen, sillä tilapäisellä kappaleen valmistamisella ei pyritä enää helpottamaan kyseisen aineiston vaan erilaisen aineiston käyttöä.”⁹⁸

Hahmotettaessa tekstin- ja tiedonlouhinta kokonaisuutena, jonka tarkoitus voi olla joko kaupallinen tai ei-kaupallinen, tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan poikkeussäännös on mielenkiintoinen. Säännöksen sisältämät kriteerit poikkeuksen soveltumiselle ja niitä täsmentävät tuomioistuinratkaisut antavat hyvän pohjan analyysille siitä, voisiko kaupallisessa tarkoituksessa toteutettava tekstin- ja tiedonlouhinta sijoittua tiettyjen edellytysten täytyessä poikkeuksen piiriin. Arvioinnin lähtöolettamaksi valitaan tällöin se, että tekstin- ja tiedonlouhintaprosessissa louhittavaksi suunniteltua sisältöä eriytetään ja/tai kopioidaan. Edellä esitetyn valossa huomautettakoon tässä vaiheessa myös,

⁹⁸ C-302/10 (Infopaq II), kohdat 49-53.

että mikäli sisällöstä laadittu kopio on luonteeltaan pysyvä, se ei voi nauttia poikkeussäännöksen tarjoamaa suojaa.

Poikkeussäännöksen soveltumisedellytykset ovat luonteeltaan kumulatiivisia, mikä tarkoittaa sitä, että jos niistä yksikin jää täyttymättä, säännöstä ei voida soveltaa kyseessä olevan kappaleen valmistamisen osalta.⁹⁹ *Triaille* on jakanut tekstin- ja tiedonlouhintaprosessin viiteen vaiheeseen, minkä jälkeen hän on analysoinut jokaisen vaiheen osalta ensinnäkin sitä, tapahtuuko kyseisessä vaiheessa kappaleen valmistamista ja toisaalta sitä, voidaanko tähän kappaleen valmistamiseen soveltaa tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan poikkeussäännöstä.¹⁰⁰ *Triaille* tulee analyysissään siihen johtopäähkseen, että louhintaprosessin jokaisessa vaiheessa saattaa tapahtua jopa useampien kappaleiden valmistamista louhittaviksi suunnitelluista sisällöistä. Lisäksi *Triaille* toteaa, että ainoastaan joissakin harvoissa tapauksissa voidaan arvioida, että louhintaprosessin eri vaiheissa valmistetut kappaleet voisivat nauttia poikkeussäännöksen suomasta rajoituksesta tai poikkeuksesta kappaleen valmistamista koskevaan oikeuteen.¹⁰¹

*Triail*län analyysi keskittyy tekstin- ja tiedonlouhintaan yleisesti, eikä louhinnan taustalla olevalle kaupalliselle tai ei-kaupalliselle tarkoitukselle ole tässä yhteydessä annettu painoarvoa. Lieneekin tulkittavissa, että jos poikkeussäännös on ilman louhinnan tarkoitukseen liittyvää raamitustakin hankalasti sovellettavissa tekstin- ja tiedonlouhintaan, niin kaupallinen tarkoitusperä voi tehdä säännöksen sovellettavuudesta entistään haastavampaa. Kaupallisessa tarkoituksessa toteutetun tekstin- ja tiedonlouhinnan kannalta erityisesti vaatimus kappaleen valmistamisen itsenäisen taloudellisen merkityksen puutteesta on ainakin ensikatsauksella ongelmallinen. Toisaalta on syytä huomauttaa, että kappaleiden valmistamisella on ainoastaan lopullisen louhintatyön mahdollistava funktio useissa tekstin- ja tiedonlouhintaprosesseissa. Tällöin on nähdäkseni edellä esitettyjen EUT:n luomien raamien puitteissa mahdollista myös argumentoida sen puolesta, että kappaleen valmistamisella tekstin- ja tiedonlouhintaprosessin yhteydessä ei ole itsenäistä taloudellista merkitystä. Kaiken kaikkiaan poikkeussäännöksen lukuisat soveltumisedellytykset ja edellytysten kumulatiivisuus johtavat kuitenkin siihen, ettei säännöksen voida katsoa oikeuttavan kaupallisessa tarkoituksessa toteutettavaa tekstin- ja

⁹⁹ C-302/10 (Infopaq II), kohta 26.

¹⁰⁰ *Triaille* ym. 2014, s. 45. Vaiheita ei ole tässä yhteydessä syytä esitellä tarkemmin, sillä ne myötäilevät tutkielman 2. luvussa esiteltyjä tekstin- ja tiedonlouhinnan perusvaiheita.

¹⁰¹ *Triaille* ym. 2014, s. 50.

tiedonlouhintaprosessia tekijänoikeudella suojatusta sisällöstä ilman tekijänoikeuden haltijan suostumusta.

Kaupallisessa tarkoituksessa suoritettavan tekstin- ja tiedonlouhinnan kannalta tietoyhteiskuntadirektiivin 5 artiklan 2 ja 3 kohtien poikkeukset tai rajoitukset ovat käytännössä merkityksettömiä. Tämä johtuu siitä, että kyseisissä kohdissa listattujen tapausten mukaisten poikkeuksien tai rajoitusten soveltumiseksi kappaleen valmistamiselta edellytetään esimerkiksi joko yksityistä käyttöä ja käyttötarkoitusta, joka ei ole välittömästi tai välillisesti kaupallinen (2 kohdan b alakohta) tai yksinomaan opetuksen tai tieteellisen tutkimuksen havainnollistamiseen liittyvää käyttöä, jolla tavoitellaan ei-kaupallista tarkoitusta (3 kohdan a alakohta).

Tässä yhteydessä on syytä myös korostaa, että vaikka tietoyhteiskuntadirektiivin 5 artiklan 2 ja 3 kohtien implementoimista ei ole säädetty jäsenvaltioille pakolliseksi, jäsenvaltioiden on täytynyt kohtien alakohtia vapaaehtoisesti implementoidessaan kuitenkin noudattaa niissä asetettuja raameja. Näin ollen, mikäli tarkoitukseltaan kaupallisessa tekstin- ja tiedonlouhintaprosessissa olisi tapahtunut kappaleen valmistamista, sillä, onko kyseessä oleva jäsenvaltio implementoinut jommankumman mainituista direktiivin 5 artiklan kohdista ei ole juuri painoarvoa. Tämä johtuu siitä, että vaikka säännös olisi implementoitu, poikkeus ei ulottuisi kaupallisessa tarkoituksessa suoritettuun tekstin- ja tiedonlouhintaan. Toisaalta mikäli säännöksiä ei olisi implementoitu lainkaan osaksi kansallista lainsäädäntöä, tekstin- ja tiedonlouhinnan mahdollistavaa poikkeusta ei direktiivin pohjalta tulkittuna olisi olemassa edes ei-kaupallista tarkoitusta ajavan louhinnan osalta.

Puhuttaessa tietoyhteiskuntadirektiivin 5 artiklan asettamista poikkeuksista ja rajoituksista saman direktiivin 2 artiklan sääntelemään kappaleen valmistamista koskevaan oikeuteen ei voida jättää huomiotta 5 artiklan 5 kohtaa, joka sisältää aiemmin esitellyn kolmivaihetestin. Nähdäkseni kolmivaihetesti kuitenkin tekee tietoyhteiskuntadirektiivistä yhä tiukemman suhteessa kaupallisessa tarkoituksessa toteutettavaan tekstin- ja tiedonlouhintaan. Etenkin toinen edellytys, jossa poikkeuksen tai rajoituksen mahdollistamalta teoksen käytöltä vaaditaan ristiriidattomuutta teoksen tavanomaiseen hyödyntämiseen nähden vaikuttaa tekstin- ja tiedonlouhinnan kannalta ongelmalliselta, sillä louhinnan voitaneen olettaa useissa tapauksissa sijoittuvan teosten tavanomaisen käytön ulkopuolelle.

3.1.2 Tietokantadirektiivi

Tietokantadirektiivi käsittelee tietokantojen oikeudellista suojaamista tekijänoikeussääntelyn lisäksi myös niin kutsutun sui generis -oikeuden kautta. Ensiksi mainitulta kantilta tietokantojen suojaan pureudutaan direktiivin II luvussa, kun taas jälkimmäinen suojamuoto saa vuoronsa III luvussa. *Schneiderin* mukaan tietokantadirektiivi oli säätämisen aikaan erityisen merkittävä kahdesta syystä. Ensinnäkin direktiivi tarjoaa tietokantojen luovalle valinnalle ja järjestämiselle yhtenäisen tekijänoikeussuojan läpi EU:n jäsenvaltioiden. Toiseksi avainseikaksi Schneider nostaa sen, että direktiivissä luotiin sui generis -oikeus, jonka tietokannan valmistajalle antama kiello-oikeus perustuu tekijänoikeudellisten näkökulmien sijaan tietokantaan liittyvien investointien suuruuteen.¹⁰²

Tietokantadirektiivi sisältää kaksi tekstin- ja tiedonlouhinnan kannalta merkityksellistä säännöstä. Direktiivin 5 artiklan a alakohdan mukaan tietokannan tekijällä on yksinoikeus kyseisen tietokannan tekijänoikeudella suojattavan ilmaisumuodon suhteen suorittaa muun muassa seuraavat toimet tai antaa lupa niihin: tietokannan tai sen osan tilapäinen tai pysyvä toisintaminen millä tavoin ja missä muodossa tahansa. Saman direktiivin 7 artiklan 1 kohdan mukaan jäsenvaltioiden on säädettävä tietokannan valmistajalle oikeus kieltää sellaisen tietokannan, jonka sisällön kerääminen, varmistaminen tai esittäminen merkitsee määrällisesti ja/tai laadullisesti huomattavaa investointia, koko sisällön tai laadullisesti ja/tai määrällisesti arvioituna olennaisen osan kopiointi ja/tai uudelleenkäyttö. Lisättäköön vielä, että saman direktiivin 7 artiklan 5 kohdan mukaan tietokannan sisällön epäolennaisien osien jatkuva ja järjestelmällinen kopiointi tai uudelleenkäyttö, jotka voisivat johtaa tämän tietokannan tavanmukaisen käytön vastaisiin toimiin tai jotka voisivat aiheuttaa kohtuutonta vahinkoa tietokannan valmistajan laillisille eduille, eivät ole sallittuja.

Esitellyt tietokantadirektiivin artikkelit konkretisoivat käsillä olevan alaluvun ensimmäisessä kappaleessa sanottua. Siinä missä 5 artiklan a alakohdassa säännellään teoskynnyksen ylittävän tietokannan tekijän tekijänoikeudellista yksinoikeutta suorittaa teoksensa suhteen tiettyjä toimia tai antaa lupa niihin, 7 artiklan 1 kohdassa onkin kyse tekijänoikeudellisesta arvioinnista irrotetusta sui generis -oikeudesta. 7 artiklan 5 kohta toimii puolestaan eräänlaisena täsmennyksenä 1 kohdan asettamille vaatimuksille. Tekstin- ja tiedonlouhinnan suorittamisen näkökulmasta näiden erilaisten tietokantasuojan muotojen syvärakenteisiin ei ole tämän esityksen puitteissa syytä pureutua tarkemmin. Riittäneekin tunnistaa, että

¹⁰² Schneider 1998, s. 551.

tietokanta, joka ei nauti tekijänoikeussuojaa voi silti saada suojaa sui generis -oikeuden perusteella.

Molempien tietokantadirektiiviin sisällytettyjen suojamuotojen tekijälleen luomista yksinoikeuksista voidaan direktiivin nojalla myöskin poiketa. Tekijänoikeuteen nojaavan tietokantasuojan osalta tällainen rajoitussäännös on direktiivin 6 artikla, jonka 2 kohdan mukaan jäsenvaltiot voivat säätää 5 artiklassa tarkoitettuihin oikeuksiin rajoituksia seuraavissa tapauksissa: a) muussa kuin elektronisessa muodossa olevan tietokannan toisintaminen yksityisiin tarkoituksiin; b) käyttö yksinomaan opetuksen tai tieteellisen tutkimuksen havainnollistamiseen, siinä määrin, kun ei-kaupalliseen tavoitteeseen pyrkiminen siihen oikeuttaa ja edellyttäen, että lähde mainitaan aina; c) käyttö yleisen turvallisuuden varmistamiseksi tai hallinnollista tai oikeudellista menettelyä varten; d) niiden kansallisessa oikeudessa perinteisesti säädettyt muut tekijänoikeutta koskevat rajoitukset, sanotun kuitenkin rajoittamatta a, b ja c alakohdan säännösten soveltamista. Sui generis -oikeuden nojalla suojaa nauttivan tietokannan valmistajalleen kielto-oikeuksia tuottavasta lähtökohdasta voidaan puolestaan poiketa tietokantadirektiivin 9 artiklan nojalla. Artiklassa nimetyt poikkeukset sui generis -oikeuteen vastaavat edellä esitellyn 6 artiklan 2 kohdan alakohdissa a, b ja c mainittuja skenaarioita.

Edellä esiteltyt säännökset asettivat tietokannoista suoritettavan tekstin- ja tiedonlouhinnan raamit ennen DSM-direktiivin voimaantuloa. Raamit olivat melko tiukat ottaen huomioon, että tekstin- ja tiedonlouhinnan toteuttaminen vaatii usein tiettyjä esitoimenpiteitä, jotka voivat pitää sisällään esimerkiksi louhittavaksi suunnitellun sisällön eriyttämistä ja/tai kopiointia.¹⁰³ Tällaisten esitoimenpiteiden suorittaminen voitaneen melko helposti todeta loukkaavan sekä tietokantadirektiivin 5 artiklan a alakohdan mukaista tietokannan tekijän yksinoikeutta tietokannan tai sen osan tilapäiseen tai pysyvään toisintamiseen että saman direktiivin 7 artiklan 1 kohdan mukaista tietokannan valmistajan oikeutta kieltää tietokannan koko sisällön tai laadullisesti ja/tai määrällisesti arvioituna olennaisen osan kopiointi ja/tai uudelleenkäyttö. Tiettyjä louhintamenetelmiä käytettäessä on kuitenkin niin, että tekstin- ja tiedonlouhintaprosessin toteuttaminen ei vaadi louhittavaksi suunnitellun sisällön eriyttämistä ja/tai kopiointia. Lisäksi tilapäisen kopioinnin osalta on huomioitava myös tietoyhteiskuntadirektiivin kappaleen valmistamista koskevan oikeuden sääntely, joka voi sanotun säädöksen 5 artiklan 1 kohdan nojalla antaa mahdollisuuden tekstin- ja

¹⁰³ Rosati 2019, s. 10.

tiedonlouhintaan tekijänoikeudella suojatusta tietokannasta ilman oikeudenhaltijan hyväksyntää.¹⁰⁴

Tietokantadirektiivin 6 ja 9 artiklojen asettamat poikkeukset löyhentävät tietokannoista suoritettavan tekstin- ja tiedonlouhinnan juridista rajoittuneisuutta. Artiklojen perusteella mahdollistetaan sellainenkin tekstin- ja tiedonlouhintaprosessi, jonka osaksi kuuluu esitoimenpiteitä, jotka sisältävät tiedon poimintaa sisällöstä ja/tai sisällön kopioimista. Artiklojen tutkielman viitekehyksessä merkitykselliset poikkeukset sisältyvät 6 artiklan 2 kohdan b alakohtaan ja 9 artiklan b alakohtaan. Alakohtien vaatimusten täyttyminen edellyttää kuitenkin kummassakin tapauksessa tekstin- ja tiedonlouhinnan ei-kaupallista päämäärää ja kopioimisen taustalla olevaa opetuksen tai tieteellisen tutkimuksen havainnollistamisen funktiota. Näin ollen niistä ei ole oikeuttamaan kaupallisessa tarkoituksessa toteutettavaa tekstin- ja tiedonlouhintaa.

3.1.3 Tietokoneohjelmadirektiivi

Tietokoneohjelmadirektiivi on annettu huhtikuussa 2009, mutta sen juuret ulottuvat huomattavasti tätä aiempaan aikaan. Tietokoneohjelmadirektiivin johdanto-osan 1 perustelukappaleen mukaan nimittäin tietokoneohjelmien oikeudellisesta suojasta vuonna 1991 annetun neuvoston direktiivin 91/250/ETY sisältöä muutettiin ajan saatossa, minkä seurauksena muutoksia kokenut direktiivi haluttiin selkeyden ja järjeistämisen vuoksi kodifioida uudelleen tietokoneohjelmadirektiivin muodossa.

Tietokoneohjelmadirektiivin 4 artiklassa säännellään tietokoneohjelman tekijän tekijänoikeudesta kumpuavia yksinoikeuksia. Artiklan 1 kohdan a ja b alakohtien mukaan jollei 5 tai 6 artiklasta muuta johdu, 2 artiklassa tarkoitetun oikeudenhaltijan yksinoikeudet käsittävät oikeuden muun muassa seuraaviin toimiin tai niiden sallimiseen: a) tietokoneohjelman tai sen osan pysyvä tai tilapäinen toisintaminen millä tavoin ja missä muodossa tahansa. Jos tietokoneohjelman lukeminen tietokoneen muistiin, näyttäminen, ajaminen, siirtäminen tai tallentaminen edellyttää tällaista toisintamista, nämä toimet edellyttävät oikeudenhaltijan suostumusta; b) tietokoneohjelman kääntäminen, muuntaminen, sovittaminen ja muunlainen muuttaminen sekä näiden toimien tulosten toisintaminen rajoittamatta kuitenkaan ohjelmaa muuttaneen henkilön oikeuksia.

¹⁰⁴ Rosati 2019, s. 10.

Tietokoneohjelmadirektiivin 5 artiklassa asetetaan poikkeus edellä esiteltyihin oikeudenhaltijan yksinoikeuksiin. Artiklan 1 kohdan mukaan jollei toisin ole nimenomaisesti sovittu, 4 artiklan 1 kohdan a ja b alakohdassa tarkoitettut toimet eivät edellytä oikeudenhaltijan suostumusta, jos ne ovat tarpeen, jotta tietokoneohjelman laillisesti hankkinut henkilö voi käyttää ohjelmaa aiotun tarkoituksen mukaisesti, virheiden korjaaminen mukaan lukien.

Vaikuttaa selvältä, että tietokoneohjelmadirektiivi ei anna tietokoneohjelman käyttäjälle mahdollisuutta tekstin- ja tiedonlouhinnan toteuttamiseen ohjelmasta. Asia voisi toki olla toisinkin, mikäli direktiivin 5 artiklan 1 kohdassa mainittu aiottu tarkoitus pitäisi terminä sisällään myös tekstin- ja tiedonlouhinnan. Tietokoneohjelmadirektiivin 10 perustelukappaleessa kuitenkin määritellään tietokoneohjelman tarkoitukseksi ”kommunikoida ja toimia yhdessä tietokonejärjestelmän muiden komponenttien sekä käyttäjien kanssa”. Lieneekin perusteltua olettaa, että tekstin- ja tiedonlouhinnan toteuttaminen tietokoneohjelmasta on toimenpiteenä arvioitu juridisesti vasta DSM-direktiiviin sisällytetyn viittauksen kautta.

3.2 DSM-direktiiviä seuraava oikeustila

3.2.1 DSM-strategiasta tekijänoikeuden ja tekstin- ja tiedonlouhinnan kannalta

DSM-strategia on Jean-Claude Junckerin johtaman EU:n komission vuonna 2015 julkaisema ohjelma, johon sisältyy digitaalisiin sisämarkkinoihin liittyviä sääntelyhankkeita ja toimenpideohjelmia. Strategialla oli jo sen valmisteluvaiheessa merkittävä painoarvo, sillä EU:n komissio nimesi strategian viimeistelyn yhdeksi kymmenestä prioriteetistaan vuonna 2014.¹⁰⁵ Strategia on rakennettu kolmen pilarin varaan: 1. Kuluttajien ja yritysten verkkotuotteiden ja -palvelujen saantia parannetaan koko Euroopassa; 2. Luodaan digitaaliverkoille ja -palveluille suotuisat olosuhteet; 3. Maksimoidaan Euroopan digitaalitalouden kasvupotentiaali.¹⁰⁶ Jokainen pilari koostuu eräänlaisista alatavoitteista, jotka täsmentävät niitä keinoja, joiden avulla pilarien määrittelemiä päämääriä tulisi tavoitella.

DSM-strategian tekijänoikeuteen liittyvä tavoite sijaitsee ensimmäisen pilarin alla, ja se on nimetty seuraavasti: ”Digitaalisen sisällön saannin parantaminen – nykyaikainen, eurooppalaisempi tekijänoikeusjärjestelmä”. Tavoitetta on taustoitettu myös tekstin- ja

¹⁰⁵ Renda 2017, s. 197.

¹⁰⁶ Digitaalisten sisämarkkinoiden strategia Euroopalle, COM(2015), s. 3-4.

tiedonlouhinnan näkökulmasta: ”Tekstin ja datan louhintaan (eli tekstin ja datan kopioiminen merkityksellisten korrelaatioiden tai esiintymien löytämiseksi) perustuvaan tutkimukseen liittyvää sekä kaupallista että epäkaupallista innovointia saattavat haitata epäselvä oikeudellinen kehys ja erilaiset kansalliset lähestymistavat. Saattaa olla tarpeen lisätä oikeusvarmuutta, jotta tutkijat ja oppilaitokset voisivat hyödyntää tekijänoikeussuojattua materiaalia laajemmin, myös rajojen yli, ja jotta ne voisivat hyötyä tekniikan tarjoamista mahdollisuuksista ja rajat ylittävästä yhteistyöstä. Samoin kuin muidenkin tekijänoikeuksia koskevien ehdotusten kohdalla on arvioitava toimenpiteen vaikutukset kaikkiin asianomaisiin osapuoliin”.¹⁰⁷

Lainaus komission tiedonannosta ei täysimääräisesti taustoita niitä syitä, joiden vuoksi tekijänoikeusjärjestelmää sekä tekstin- ja tiedonlouhinnan sääntelyä järjestelmän osana päätettiin lähteä muokkaamaan digitalisoituvan yhteiskunnan tarpeita vastaaviksi. Lainauksesta voidaan kuitenkin erottaa kaksi epäkohtaa, joiden korjaamista voitaneen pitää tärkeinä päämäärinä tekstin- ja tiedonlouhintaa koskevan sääntelyuudistuksen taustalla. Nämä epäkohdat ovat tekstin- ja tiedonlouhinnan epäselvä oikeudellinen kehys sekä erilaiset kansalliset lähestymistavat. Lainauksen perusteella on edelleen pääteltävissä, että epäkohtien korjaamisen taustalla vaikutti pyrkimys oikeusvarmuuden kohentamisesta kyseisessä asiakokonaisuudessa.

Tekstin- ja tiedonlouhinnan oikeudellisen kehyksen epäselvyys käy ilmi jo tutkielmassa aiemmin esiteltyjä tietoyhteiskuntadirektiivin, tietokantadirektiivin ja tietokoneohjelmadirektiivin artikloita tarkastellessa. Kuten havaittua, kahteen ensiksi mainittuun direktiiviin on sisällytetty tekstin- ja tiedonlouhinnan ainakin teoreettisesti mahdollistavia poikkeussäännöksiä. Poikkeussäännökset kuitenkin rajautuvat kategorisesti tiettyjen reunaehtojen täyttymiseen, kuten esimerkiksi kopioimiseen opetuksen tai tieteellisen tutkimuksen havainnollistamiseksi tai kopioimisen ei-kaupalliseen tarkoitukseen. Näin ollen siinä missä ei-kaupallisessa tarkoituksessa toteutettu tekstin- ja tiedonlouhinta jää direktiivien valossa oikeudellisen hyväksyttävyytensä osalta hieman epäselväksi, on kaupallisessa tarkoituksessa toteutettu louhinta samojen artiklojen nojalla käytännössä mahdotonta. Tätä taustaa vasten onkin ymmärrettävää, että komission tiedonannossa on mainittu myös tekstin- ja tiedonlouhintaan liittyvän kaupallisen innovoinnin mahdollinen heikentyminen epäselvän oikeudellisen kehyksen seurauksena. Tiedonannossa on kuitenkin

¹⁰⁷ Digitaalisten sisämarkkinoiden strategia Euroopalle, COM(2015), s. 7.

painotettu selvästi enemmän tutkijoiden ja oppilaitosten tarvetta sääntelyuudistuksen tuomalle mahdollisuudelle hyödyntää tekijänoikeussuojattua materiaalia laajemmin. Painotus tuli lopulta näkymään myös DSM-direktiivin lopullisessa versiossa, jonka sisältöön pureudutaan tarkemmin seuraavassa luvussa.

Erilaisiin kansallisiin lähestymistapoihin liittyvällä epäkohdalla viitattaneen komission tiedonannossa nimenomaan tietoyhteiskuntadirektiivin ja tietokantadirektiivin poikkeussäännöksiin, joiden implementoimista ei ole säädetty jäsenvaltioille pakolliseksi. Tällöin tarkoitetaan tietoyhteiskuntadirektiivin 5 artiklan 2 ja 3 kohtien alakoh tia sekä tietokantadirektiivin 6 ja 9 artikloja. Ensinnäkin on todettava, että jokainen näistä säännöksistä edellyttää jäsenvaltion niitä soveltaessaankin sellaisten opetuksen tai tieteellisen tutkimuksen havainnollistamiseen ja ei-kaupalliseen tarkoitukseen liittyvien vaatimusten täyttymistä, että kaupallisessa tarkoituksessa toteutettu tekstin- ja tiedonlouhintaa jää niiden soveltamisalan ulkopuolelle. Mitä taas tulee ei-kaupallisessa tarkoituksessa toteutettuun tekstin- ja tiedonlouhintaan, poikkeussäännösten soveltuminen ja sitä kautta louhinnan ainakin jonkinasteinen sallittavuus ovat määräytyneet sen perusteella, ovatko jäsenvaltiot implementoineet ne osaksi kansallista lainsäädäntöään ja jos ovat, miten ne ovat sen tehneet. Tämän seurauksena kansalliset lähestymistavat ovat eriytyneet toisistaan ei-kaupallisen tekstin- ja tiedonlouhinnan suhteen.

Esitellyt epäkohdat luovat nähdäkseni hyvän perustan tekstin- ja tiedonlouhinnan sääntelykehyksen uudistamiselle. Tästä huolimatta jo komission tiedonannosta otetun lainauksen perusteella jää epäily siitä, miten kaupallisessa tarkoituksessa toteutetun tekstin- ja tiedonlouhinnan oikeudellista kehystä voitaisiin selkeyttää. Seuraavassa luvussa pureudutaankin tarkemmin DSM-direktiivin 4 artiklaan, jonka avulla tämä päämäärä on pyritty saavuttamaan.

3.2.2 DSM-direktiivin 4 artiklan sisältö ja perustelut

DSM-direktiivin 4 artikla sisältää tekstin- ja tiedonlouhintaa koskevan rajoituksen tai poikkeuksen. Artiklan 1 kohdan mukaan jäsenvaltioiden on säädettävä poikkeuksesta tai rajoituksesta direktiivin 96/9/EY 5 artiklan a alakohdassa ja 7 artiklan 1 kohdassa, direktiivin 2001/29/EY 2 artiklassa, direktiivin 2009/24/EY 4 artiklan 1 kohdan a ja b alakohdassa sekä tämän direktiivin 15 artiklan 1 kohdassa säädettyihin oikeuksiin, kun kyseessä on kappaleen valmistaminen ja kopiointi tekstin- ja tiedonlouhintaa varten teoksista ja muusta suojatusta aineistosta, johon on laillinen pääsy. 4 artiklan 2 kohdan mukaan 1 kohdan mukaisesti

valmistettuja kappaleita ja kopioita voidaan säilyttää niin kauan, kuin tekstin- ja tiedonlouhintaa varten on tarpeen. Edelleen saman artiklan 3 kohdan mukaan 1 kohdassa säädettyä poikkeusta tai rajoitusta sovelletaan edellyttäen, että oikeudenhaltijat eivät ole nimenomaisesti pidättäneet oikeuksiaan kyseisessä kohdassa tarkoitettuun teosten ja muun suojatun aineiston käyttöön asianmukaisella tavalla, kuten koneluettavilla keinoilla, kun kyseessä on verkossa yleisön saataviin saatettu sisältö. Artiklan 4 kohdan mukaan artikla ei vaikuta direktiivin 3 artiklan soveltamiseen.

DSM-direktiivin johdanto-osan 18 perustelukappale koskee erityisesti 4 artiklaa. Perustelukappaleessa tasapainoillaan oikeudenhaltijoiden ja tekijänoikeudella suojattua sisältöä tekstin- ja tiedonlouhinnassa hyödyntävien tahojen etunäkökohtien välillä. Yhtäältä perustelukappaleessa painotetaan oikeudenhaltijoiden oikeutta lisensoida DSM-direktiivin 3 artiklassa säädetyn, tekstin- ja tiedonlouhintaa tieteellisiin tarkoituksiin oikeuttavan poikkeussäännöksen ja tietoyhteiskuntadirektiivissä säädettyjen poikkeusten ja rajoitusten ulkopuolelle jäävän suojatun aineistonsa käyttö. Toisaalta kappaleessa kuitenkin tunnistetaan sisältöä tekstin- ja tiedonlouhintaan käyttävien tahojen oikeusvarmuuden tarve etenkin silloin, kun louhintaprosessiin ei sen luonteesta johtuen voida soveltaa tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan mukaista poikkeussäännöstä. Nimenomaan oikeusvarmuusaspektia ja yksityisen sektorin innovoinnin kannustamista painottaen perustelukappaleessa todetaan, että ”tällä direktiivillä olisi säädettävä tietyin edellytyksin poikkeuksesta tai rajoituksesta, joka koskee kappaleen valmistamista ja kopiointia teoksista tai muusta suojatusta aineistosta tekstin- ja tiedonlouhintaa varten, ja sallittava valmistettujen kopioiden säilyttäminen, niin kauan kuin se on tarpeen tekstin- ja tiedonlouhintaa varten”.

DSM-direktiivin johdanto-osan 18 perustelukappaleessa mainitaan 4 artiklaankin sisällytetyt edellytykset poikkeuksen soveltamiseen. Edellytykset koskevat laillista pääsyä suojattuun sisältöön ja oikeudenhaltijoiden mahdollisuutta pidättää asianmukaisella tavalla oikeutensa kappaleen valmistamiseen ja kopiointiin tekstin- ja tiedonlouhintaa varten. Ensiksi mainittua edellytystä ei täsmennetä 18 perustelukappaleessa, mutta 14 perustelukappaleessa laillinen pääsy määritellään seuraavasti: ”Laillinen pääsy olisi ymmärrettävä siten, että se kattaa pääsyn sisältöön avointa saatavuutta koskevien toimintaperiaatteiden (open access policy) perusteella taikka oikeudenhaltijoiden ja tutkimusorganisaatioiden tai kulttuuriperintölaitosten sopimusjärjestelyjen, kuten tilausten,

perusteella taikka muilla laillisilla keinoilla. Tapauksissa, joissa tilauksia ovat tehneet tutkimusorganisaatiot tai kulttuuriperintölaitokset, olisi esimerkiksi katsottava, että laillinen pääsy on niihin yhteydessä olevilla, näiden tilausten kattamilla henkilöillä. Laillisen pääsyn olisi katettava myös pääsy verkossa vapaasti saatavilla olevaan sisältöön.”

Vaikka 14 perustelukappaleessa puhutaan DSM-direktiivin 3 artiklan hengessä tutkimusorganisaatioista ja kulttuuriperintölaitoksista, voitaneen olettaa, että samat reunaehdot pätevät myös 4 artiklassa mainitun laillisen pääsyn määrittelyssä. Mikäli näin ei kuitenkaan katsottaisi olevan, *Geiger – Frosio – Bulayenko* ovat esittäneet, että laillisen pääsyn tulkinnassa voitaisiin analogisesti hyödyntää tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan poikkeussäännöksen lailliselle käytölle antamaa merkityssisältöä.¹⁰⁸ Merkityssisältö on käytännössä annettu tietoyhteiskuntadirektiivin johdanto-osan 33 perustelukappaleessa, jonka mukaan ”käyttöä olisi pidettävä laillisena, jos se on oikeudenhaltijan sallimaa tai jos sitä ei ole lailla rajoitettu”.

Jälkimmäistä, oikeudenhaltijan oikeuksien pidättämistä koskevaa edellytystä täsmennetään puolestaan DSM-direktiivin johdanto-osan 18 perustelukappaleessa seuraavalla tavalla: ”Kun sisältö on saatettu yleisön saataviin verkossa, asianmukaiseksi olisi katsottava kyseisten oikeuksien pidättäminen ainoastaan koneluettavilla keinoilla, mukaan lukien metatiedot ja verkkosivuston tai palvelun ehdot. Oikeuksien pidättäminen tekstin- ja tiedonlouhintaa varten ei saisi vaikuttaa muuhun käyttöön. Muissa tapauksissa voi olla asianmukaista pidättää oikeudet muilla keinoilla, kuten sopimusjärjestelyillä tai yksipuolisella ilmoituksella. Oikeudenhaltijoiden olisi voitava soveltaa toimenpiteitä varmistaakseen, että niiden tätä koskevia oikeuksien pidättämisiä kunnioitetaan”.

Myös DSM-direktiivin johdanto-osan 9 perustelukappaleeseen sisältyy direktiivin 4 artiklan tulkinnan kannalta merkityksellisiä havaintoja. Perustelukappaleessa todetaan, että ”tekstin- ja tiedonlouhintaa voidaan harjoittaa myös pelkistä tosiasioista tai tiedoista, jotka eivät ole tekijänoikeudella suojattuja, eikä tällaisissa tapauksissa tekijänoikeuslainsäädännön mukaista lupaa tarvita”. Lisäksi perustelukappaleessa huomioidaan tilanteet, joissa tekstin- ja tiedonlouhintaan ei sisälly kappaleen valmistamista tai joihin voidaan kappaleen valmistamisen tilapäisyyden ja muiden edellytysten täyttymisen perusteella soveltaa tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan poikkeussäännöstä.

¹⁰⁸ *Geiger – Frosio – Bulayenko* 2018, s. 32.

3.2.3 Tekstin- ja tiedonlouhinta kaupallisessa tarkoituksessa: Mikä muuttuu?

Jotta voitaisiin analysoida, mitä muutoksia DSM-direktiivi tuo kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan sääntelylliseen kehykseen, on syytä käydä ensiksi lyhyesti läpi, miten ei-kaupallisessa ja tieteellistä tutkimusta edistävässä tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan kenttä tulee muuttumaan DSM-direktiivin 3 artiklan myötä.

Lyhyt vilkaisu DSM-direktiivin 3 artiklaan on nähdäkseni perusteltua kahdesta syystä. Ensinnäkin edellä esitelty tietoyhteiskuntadirektiivin ja tietokantadirektiivin tekstin- ja tiedonlouhintaa koskevat säännökset ja niihin säädetyt poikkeukset tai rajoitukset soveltuvat jo todetulla tavalla ennemmin ei-kaupallisessa tarkoituksessa toteutettavaan tekstin- ja tiedonlouhintaan kuin sen kaupallisen tarkoituksen sisältävään vastineeseen. Näin ollen DSM-direktiivin mukanaan tuomia muutoksia täytyy osittain arvioida nimenomaan ei-kaupallisen tarkoituksen perspektiivistä. Toiseksi kuvatus kaltainen lähestymistapa auttaa hahmottamaan sitä, onko tekstin- ja tiedonlouhinnan sääntelyä päädytty muuttamaan tasapuolisesti sekä ei-kaupalliset että kaupalliset tarkoitukset huomioon vai onko jompaakumpaa niistä painotettu toista enemmän. 3 artiklaa tullaan kuitenkin arvioimaan melko rajatusti, sillä ei-kaupallisessa tarkoituksessa toteutettava tekstin- ja tiedonlouhinta ei tietyistä relevanttiudesta huolimatta lukeudu tutkielman pääasiallisten kiinnostuksen kohteiden joukkoon.

Kuten edellä on havaittu, tietoyhteiskuntadirektiivin ja tietokantadirektiivin tekstin- ja tiedonlouhinnan kannalta olennaiset poikkeukset perustuvat pitkälti opetuksen tai tieteellisen tutkimuksen havainnollistamiseen liittyvään sisällön käyttöön ja ylipäänsä ei-kaupallisessa tarkoituksessa toimimiseen. Tällaisten poikkeussäännösten implementoiminen on jo todetulla tavalla jätetty jäsenvaltioiden oman harkinnan varaan.

Ei-kaupallisen tekstin- ja tiedonlouhinnan oikeudellista kehystä on pyritty selkiyttämään ja yhtenäistämään DSM-direktiivin 3 artiklan avulla. 3 artiklan 1 kohdan mukaan jäsenvaltioiden on säädettävä poikkeuksesta direktiivin 96/9/EY/ 5 artiklan a alakohdassa ja 7 artiklan 1 kohdassa, direktiivin 2001/29/EY 2 artiklassa ja DSM-direktiivin 15 artiklan 1 kohdassa säädettyihin oikeuksiin, kun kyseessä on kappaleen valmistaminen ja kopiointi, joita tutkimusorganisaatiot ja kulttuuriperintölaitokset toteuttavat suorittaakseen tieteellistä tutkimusta varten tekstin- ja tiedonlouhintaa teoksista tai muusta suojatusta aineistosta, joihin niillä on laillinen pääsy. Artiklan 2, 3 ja 4 kohdissa säädetään valmistettujen

kappaleiden ja kopioiden säilyttämisestä, oikeudenhaltijoille mahdollistetuista toimenpiteistä sekä parhaiden käytäntöjen määrittelemisestä 2 ja 3 kohtia sovellettaessa.

Tarttumatta DSM-direktiivin 3 artiklan epäkohtiin tutkimusorganisaatioiden ja kulttuuriperintölaitosten näkökulmasta ¹⁰⁹ voidaan huomata, että artikla suhtautuu vähemmän rajoittavasti tekstin- ja tiedonlouhintaan kuin sen kaupallisessa tarkoituksessa toteutettavaa tekstin- ja tiedonlouhintaa koskeva vastinsäännös direktiivin 4 artiklassa.¹¹⁰ 3 artikla asettaa jäsenvaltioille pakollisesti implementoitavan velvollisuuden säätää poikkeuksesta tietoyhteiskuntadirektiivin kappaleen valmistamista koskevaan yksinoikeuteen ja tietokantadirektiivin mukaiseen tekijänoikeuteen tai sui generis -oikeuteen perustuvaan yksinoikeuteen. Tämä on toteutettu tavalla, joka ei anna oikeudenhaltijalle mahdollisuutta pidättää oikeutta kappaleen valmistamiseen tekijänoikeudella suojatusta teoksestaan tai muusta sisällöstä.¹¹¹

Peilattaessa 3 artiklaa edellä käsiteltyyn komission tiedonantoon ja siinä huomioituihin epäkohtiin on selvää, että artiklassa on tartuttu melko hyvin ei-kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan epäselvään oikeudelliseen kehykseen ja erilaisiin kansallisiin lähestymistapoihin. Oikeudellisen kehyksen voitaneen olettaa selkiytyvän, kun poikkeukset mainittuihin tietoyhteiskuntadirektiivin ja tietokantadirektiivin suomiin yksinoikeuksiin eivät ole hajautetusti eri säännöksissä vailla mainintaa itse tekstin- ja tiedonlouhinnasta, vaan siitä on 3 artiklan kautta säädetty suoraan kyseisiin yksinoikeuksiin viitaten. Erilaiset kansalliset lähestymistavat eivät puolestaan tulle nousemaan asiakokonaisuuden kannalta enää ongelmallisiksi, sillä aiemmasta lähestymistavasta poiketen DSM-direktiivissä poikkeussäännös on säädetty jäsenmaiden pakollisesti implementoitavaksi.

Palatkaamme taas DSM-direktiivin 4 artiklaan ja siihen, miten kyseinen säännös muuttaa kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan tekijänoikeudellista sääntelykehystä. *Caspers – Guibault* ovat katsoneet, että tietoyhteiskuntadirektiivissä kappaleen valmistamista koskevalle oikeudelle asetetut poikkeukset tai rajoitukset ovat

¹⁰⁹ Ks. pääasiallisista epäkohdista kuitenkin esim. Ducato – Strowel 2019.

¹¹⁰ Ks. kuitenkin esim. Quintais 2020, s. 32, jolla todetaan molempien artikloiden olevan laajuudeltaan kapeita ja sisältävän niiden täysimääräistä soveltamista estäviä rajoituksia.

¹¹¹ Huomautettakoon, että 3 artiklan 3 kohdan mukaan oikeudenhaltijoiden on voitava soveltaa toimenpiteitä, joilla varmistetaan niiden verkkojen ja tietokantojen turvallisuus ja eheys, joissa teoksia tai muuta suojattua aineistoa säilytetään. Tällaiset toimenpiteet eivät saa ylittää sitä, mikä on tarpeen tämän tavoitteen saavuttamiseksi.

vaikuttaneet olleen liian rajoittuneita kattaakseen tekstin- ja tiedonlouhintaprosessiin olennaisesti kuuluvia toimenpiteitä.¹¹² Edelleen he ovat todenneet, että niissäkin tapauksissa, joissa kyseiset poikkeukset tai rajoitukset on katsottu soveltuviksi tekstin- ja tiedonlouhintaan, ne ovat rajoittuneet ainoastaan pieneen määrään tarkoituksia ja edunsaajia.¹¹³

Edellä esitellyt näkemykset ovat perusteltuja etenkin kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan kannalta, sillä tietoyhteiskuntadirektiivin poikkeussäännökset edellyttävät soveltuakseen nimenomaan ei-kaupallista tarkoitusta ja tietyiltä osin myös opetuksen tai tieteellisen tutkimuksen havainnollistamiseen liittyvää tavoitetta. Samojen lähtökohtien voidaan katsoa koskevan myös tietokantadirektiivin tekstin- ja tiedonlouhintaa koskevia poikkeussäännöksiä. Näin ollen tietoyhteiskuntadirektiivin ja tietokantadirektiivin asettamat raamit ovat luoneet kaupallisessa tarkoituksessa toteutettavalle tekstin- ja tiedonlouhinnalle melko hajanaisen perustan, jonka pohjalta on vaikeaa hahmottaa, millaisissa tilanteissa kuvatus kaltaisen tekstin- ja tiedonlouhinta olisi lainsäädännöllisesti sallittua.¹¹⁴

Hugenholtz on todennut, että DSM-direktiivin 3 ja 4 artiklat antavat tekstin- ja tiedonlouhinnalle huomattavasti vähemmän vapauksia kuin ensikatsomalta saattaa vaikuttaa.¹¹⁵ Erityisesti 4 artiklan osalta *Rosati* on katsonut, että oikeudenhaltijoille jätetty mahdollisuus pidättää oikeutensa kappaleen valmistamiseen teoksestaan tai muusta suojatusta aineistostaan muun muassa tekstin- ja tiedonlouhintaan liittyvässä tarkoituksessa saattaa käytännössä heikentää mahdollisuuksia soveltaa artiklaa tai jopa kadottaa sen merkityksen kokonaisuudessaan.¹¹⁶ Näkemykset vaikuttavat perustelluilta, mutta ennen niiden allekirjoittamista on syytä käydä 4 artiklan mukanaan tuomat muutokset kokonaisvaltaisemmin läpi. Tämän jälkeen voidaan paremmin analysoida, tekeekö artiklan 3 kohta nämä muutokset käytännössä tyhjiksi.

Kuten edellä on todettu, DSM-direktiiviä edeltäneessä oikeustilassa kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan käytännössä ainoa

¹¹² Caspers – Guibault 2016, s. 99.

¹¹³ Caspers – Guibault 2016, s. 99.

¹¹⁴ Arvioissa ei ole luonnollisesti otettu huomioon tilannetta, jossa oikeudenhaltija on antanut ulkopuoliselle taholle luvan käyttää teostaan tai muuta suojattua aineistoa kaupallisessa tarkoituksessa toteutettavaan tekstin- ja tiedonlouhintaan.

¹¹⁵ Hugenholtz 2019.

¹¹⁶ Rosati 2019, s. 21.

tekijänoikeudellisesti relevantti poikkeussäännös oli tietoyhteiskuntadirektiivin 5 artiklan 1 kohta. Tutkielmassa on myös havainnollistettu, kuinka hankalasti kyseinenkin säännös on sovellettavissa kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan sallivana poikkeuksena tai rajoituksena. Kertauksen vuoksi todettakoon, että hankaluus juontaa juurensa sekä säännöksen soveltamisedellytysten kumulatiivisesta luonteesta että joidenkin yksittäisten edellytysten asettamista vaatimuksista. On esimerkiksi esitetty, että tekstin- ja tiedonlouhinta vaatii useissa tapauksissa sellaista louhittavaksi suunnitellun sisällön kopiointia, jota ei voida pitää säännöksen edellyttämänä tilapäisenä kappaleen valmistamisena.¹¹⁷ Tämä taas johtaa säännöksen edellytysten kumulatiivisuuden vuoksi suoraan säännöksen soveltamiskelvottomuuteen.

DSM-direktiivin 4 artiklasta tuli tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan ohella ainoa tekstin- ja tiedonlouhinnan kannalta relevantti pakollinen poikkeus tietoyhteiskuntadirektiivin 2 artiklan asettamaan kappaleen valmistamista koskevaan oikeuteen. DSM-direktiivin 4 artiklalla ja tietoyhteiskuntadirektiivin 5 artiklan 1 kohdalla on kuitenkin merkittäviä eroavaisuuksia etenkin kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan osalta. Toisin kuin tietoyhteiskuntadirektiivin 5 artiklan 1 kohta, DSM-direktiivin 4 artikla käsittelee nimensä mukaisesti ainoastaan tekstin- ja tiedonlouhintaa ja vieläpä muuhun kuin tieteelliseen tutkimukseen liittyvän tarkoituksen kautta. Siinä missä kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan tekijänoikeudellinen sallittavuus oli tietoyhteiskuntadirektiivin poikkeussäännöksen nojalla hyvin epäselvä ja vaikeasti saavutettavissa, DSM-direktiivin 4 artiklan 1 kohta yksiselitteisesti sallii sen, mikäli aineistoon on laillinen pääsy.

DSM-direktiivin 4 artikla asettaa pakollisen poikkeuksen myös tietokantadirektiivin mukaiseen tekijänoikeudelliseen ja sui generis -tyyppiseen suojaan. Tietokantadirektiiviin säädetyt poikkeukset eivät ole jäsenvaltioille pakollisesti implementoitavia, eivätkä niiden soveltamisedellytykset soinnu yhteen kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan kanssa. Näin ollen pakollinen poikkeus muuttaa oikeustilaa myös näiden suojamuotojen osalta. DSM-direktiivi toimii pakollisena poikkeuksena myös tietokoneohjelmadirektiivin 4 artiklan 1 kohdan a ja b alakohtien mukaisiin yksinoikeuksiin sekä DSM-direktiivin 15 artiklassa säädettyyn lehtijulkaisujen suojaan verkkokäytössä.

¹¹⁷ Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules, s. 2.

DSM-direktiivin 4 artiklan 2 kohta takaa tekstin- ja tiedonlouhintaa kaupallisessa tarkoituksessa toteuttavalle taholle oikeuden säilyttää valmistamiaan kappaleita ja kopioita niin kauan, kuin tekstin- ja tiedonlouhintaa varten on tarpeen. Tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan pakollinen poikkeussäännös taas edellyttää soveltuakseen muun muassa kappaleen valmistamisen tilapäisyyttä ja sen väliaikaista ja satunnaista luonnetta. Voidaankin tulkita, että DSM-direktiivin 4 artikla takaa kaupallisessa tarkoituksessa toteutettavaan tekstin- ja tiedonlouhintaan kohdistetun mahdollisuuden säilyttää louhinnan suorittamiseksi valmistettuja kappaleita ja kopioita juuri sen ajan, jonka prosessi itsessään vaatii.

Edellä kuvatut DSM-direktiivin 4 artiklan tuomat muutokset kaupallisessa tarkoituksessa toteutettavaan tekstin- ja tiedonlouhintaan vaikuttaisivat selkiyttävän oikeustilaa aiemmasta melko paljonkin. Artiklan 3 kohta tuo tähän lähtökohtaan kuitenkin ristiriitaisuutta antamalla oikeudenhaltijoille mahdollisuuden pidättää oikeutensa teostensa ja muun suojatun aineistonsa käyttöön. Vaikka oikeuksien pidättämisen tuleekin olla nimenomaista ja se on toteutettava asianmukaisella tavalla, 3 kohdalla on nähdäkseni potentiaalia häivyttää koko artiklan merkitys.

DSM-direktiivissä ei määritellä tyhjentävästi tapoja, joilla oikeudenhaltijoiden on mahdollista pidättää oikeutensa direktiivin 4 artiklan 3 kohdan mukaisesti. Direktiivin johdanto-osan 18 perustelukappaleessa kuitenkin tarkennetaan verkossa yleisön saataviin saatetun sisällön osalta, että asianmukainen keino pidättää oikeudet kappaleen valmistamiseen ja kopiointiin tekstin- ja tiedonlouhintaa varten on nimenomaan oikeuksien pidättäminen koneluettavilla keinoilla, mukaan lukien metatiedot ja verkkosivuston tai palvelun ehdot. Tältä pohjalta esimerkiksi Hugenholtz on todennut, että oikeudenhaltijat pystyvät estämään teoksistaan tai muusta suojatusta aineistostaan kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan ainoastaan lisäämällä robot.txt-tyyppisiä metatietoja verkkoaineistonsa.¹¹⁸ Tällainen toiminta tarkoittaisi käytännössä robotin rajausstandardin (robots exclusion standard) käyttämistä. Robotin rajausstandardi on www-palvelimelle luotava tiedosto, joka ilmoittaa roboteille sivuston pääsykäytännöistä.¹¹⁹ Tiedosto tulee sijoittaa verkkosivuston juurihakemistoon, ja sen tulee olla nimeltään ”robots.txt”.¹²⁰ Verkkosivuston ylläpitäjä voi oikein sijoitetun ja nimetyn robotin

¹¹⁸ Hugenholtz 2019.

¹¹⁹ Koster 1994.

¹²⁰ Yang, C. – Liao, H. 2010.

rajausstandardin avulla esimerkiksi kieltää hakukoneiden hakurobotteja prosessoimasta tiettyjä verkkosivuston osia.

DSM-direktiivin johdanto-osan 18 perustelukappaleen nojalla muulla tavoin kuin verkossa yleisön saataviin saatettua sisältöä koskevat kappaleen valmistamista koskevat oikeudet voidaan asianmukaisesti pidättää muilla kuin edellisessä kappaleessa esitetyillä keinoilla. Tällaisia keinoja voivat 18 perustelukappaleen mukaan olla sopimusjärjestelyt tai yksipuoliset ilmoitukset. Lisenssisopimusten voitaneen olettaa tulevan tässä yhteydessä ajankohtaisiksi. Oletusta vahvistaa edelleen direktiivin 18 perustelukappale, jossa painotetaan oikeudenhaltijoiden mahdollisuutta lisensoida myös sellainen teoksensa käyttö, johon direktiivin 4 artiklan poikkeus tai rajoitus kohdistuu.

Erityisesti verkossa yleisön saataviin saatettuun sisältöön liittyvien oikeuksien pidättämiselle olisi nähdäkseni voitu asettaa direktiivissä tarkemmat puitteet. Direktiivin ainoa konkreettinen vaatimus on kuitenkin tällaisten oikeuksien pidättäminen koneluettavilla keinoilla. Tällöin oikeudenhaltijoiden käytännön tavat oikeuksien pidättämiseen tullevat vaihtelevaan jonkin verran ja sitä kautta mahdollisesti aiheuttamaan epäselvyyttä siitä, mikä tapa tulkitaan riittäväksi ja mikä taas ei. *Dusollier* on myös katsonut, että oikeudenhaltijoiden 4 artiklan nojalla asettamat robotin rajaustandardit saattavat tulevaisuudessa estää myös ei-kaupallisessa tarkoituksessa toimivien tutkijoiden mahdollisuuden suorittaa tekstin- ja tiedonlouhintaa kulloinkin kyseessä olevista teoksista.¹²¹ Tällöin olisi kyse tilanteesta, jossa DSM-direktiivin 4 artikla asettuisi saman direktiivin 3 artiklan täysimääräisen toteutumisen esteeksi.

Hughenoltz on katsonut, että DSM-direktiivin 4 artiklan 3 kohta johtaa erillisen tekstin- ja tiedonlouhinnan markkinan luomiseen ja legitimoimiseen. Oikeudenhaltijat, kuten esimerkiksi julkaisijat, saattavat haluta kontrolloida tätä markkinaa lisensointimekanismien tai suorien kieltojen avulla.¹²² Huoli on nähdäkseni varsin ymmärrettävä; etenkin järjestäytyneillä aloilla voitaneen olettaa käytettävän 4 artiklan 3 kohdan mukaista opt out -mekanismia aktiivisesti. Oletus perustuu lähtökohtaan siitä, että pidättämällä oikeutensa 3 kohdan mukaisesti oikeudenhaltijoiden on mahdollista saada teostensa käyttämisestä lisätuloja asettamalla kaupallisessa tarkoituksessa toteutettava tekstin- ja tiedonlouhinta luvanvaraiseksi. Tällöin yksi teos voi synnyttää oikeudenhaltijalle tuloja sekä sisältöön

¹²¹ Dusollier 2020, s. 987.

¹²² Hughenoltz 2019.

pääsyä että sisällön louhimista koskevissa vaiheissa. Asetelma näyttäytyy varsin loogisena, eikä tutkielman tarkoituksena olekaan esittää kritiikkiä tällaista oikeudenhaltijoiden potentiaalista toimintatapaa kohtaan. Kyseenalaista on sen sijaan se, kuinka suuren haasteen DSM-direktiivin 4 artiklan 3 kohta asettaa koko artiklan relevanttiudelle.

DSM-direktiivin johdanto-osan 18 perustelukappaleen pohjalta direktiivin 4 artiklan perimmäisenä tarkoitusperänä on ensinnäkin lisätä oikeusvarmuutta tilanteissa, joissa kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan ei voida katsoa täyttävän tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan asettamia edellytyksiä. Toiseksi tavoitteeksi perustelukappaleessa mainitaan kannustaminen innovointiin myös yksityisellä sektorilla. Mikäli lähdetään siitä oletuksesta, että kaikki oikeudenhaltijat hyödyntäisivät DSM-direktiivin 4 artiklan 3 kohdan opt out -mekanismia, oikeusvarmuus toki lisääntyisi, mutta innovointiin tilanne tuskin kannustaisi. Mikäli taas opt out -mekanismin soveltamisaktiivisuudessa ilmenee aste-eroja, tilanne ei ole innovointiin kannustamista koskevan tavoitteen osalta yhtä lohduton. Seuraavaksi käsitellään lyhyesti näitä erilaisia skenaarioita.

Oikeusvarmuus paranee DSM-direktiivin myötä joka tapauksessa siksi, että aiemmin hankalasti tulkittavat tietoyhteiskuntadirektiivin 5 artiklan 1 kohdan kumulatiiviset soveltamisedellytykset vaihtuvat DSM-direktiivin 4 artiklan nimenomaiseen kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan mahdollistavaan poikkeukseen tai rajoitukseen. Oikeudenhaltijoiden aktiivisuus opt out -mekanismin käyttämisessä vaikuttaa oikeusvarmuuteen nähdäkseni ainoastaan välillisesti sillä tavalla, että jotkut aineistot ovat vapaasti louhittavissa kaupallisissakin tarkoituksissa, kun taas toisten osalta oikeudenhaltijat ovat pidättäneet oikeutensa. Tässä mielessä DSM-direktiivin 4 artikla ei onnistu tuomaan absoluuttista oikeusvarmuutta, mutta ainakin lisää sitä aiempaan oikeustilaan nähden.

Yksityisen sektorin innovointiin kannustaminen toteutuu osittain jo oikeusvarmuutta lisäämällä. Opt out -mekanismin käyttöasteella on kuitenkin edellä todetulla tavalla myös varsin suurin merkitys mainitun tavoitteen saavuttamisessa. On kovin epätodennäköistä, että kaikki oikeudenhaltijat pidättäisivät oikeutensa DSM-direktiivin 4 artiklan 3 kohdan mukaisesti. Järjestäytyneet toimialat ovat kuitenkin tottuneet hyödyntämään lisensointimekanismeja oikeudenhaltijoiden eduksi, minkä vuoksi niiden voidaan edellä todetun mukaisesti olettaa hyödyntävän 3 kohdan tarjoamaa mahdollisuutta. Muilla aloilla tilanne voisi ainakin teoriassa olla myös toisenlainen. Arviointi kulminoitune

oikeudenhaltijoiden näkökulmasta kuitenkin lopulta siihen, mitä hyötyjä ne voivat saavuttaa yhtäältä käyttämällä opt out -mekanismia tai toisaalta jättämällä mekanismin hyödyntämättä.

Oikeudenhaltijan näkökulmasta opt out -mekanismin käyttämisen hyödyksi voidaan katsoa melko yksioikoisesti taloudellinen etu. Pidättämällä oikeutensa DSM-direktiivin 4 artiklan 3 kohdan mukaisesti oikeudenhaltijat voivat nimittäin edellä todetun mukaisesti hankkia lisätuloja lisensoimalla mahdollisuuden kaupallisessa tarkoituksessa toteutettavaan tekstin- ja tiedonlouhintaan teoksestaan. Samalla ne toki pitävät kiinni myös tekijänoikeuden ytimeen kuuluvasta yksinoikeudestaan valmistaa teoksestaan kappaleita.

Opt out -mekanismin käyttämättä jättämisestä oikeudenhaltijoille koituvia hyötyjä on huomattavasti haastavampaa löytää. Voisi kuitenkin ajatella yhden sellaisen olevan yksinkertaisesti yhteiskunnan kehityksen vauhdittaminen. Ajatuksen perustana on se, että tekstin- ja tiedonlouhintaprosessin lopullisena tarkoituksena on tutkielmassa aiemmin kuvatulla tavalla uuden tiedon ja oivallusten tuottaminen samalla mahdollistaen yhteiskunnan digitalisoitumista koneoppimista ja tekoälykehitystä tukien. Tällöin on usein varsin selvää, että teoksista ei valmisteta kappaleita tarkoituksessa korvata teoksia niiden omilla markkinoilla, vaan ennemminkin halutaan käyttää teoksia datamassana, josta kone oppii muodostamaan kaavoja tai korrelaatioita, joiden pohjalta voidaan tuottaa uutta tietoa. Tämä tieto voi puolestaan tarjota uusia virikkeitä myös kyseessä olevan toimialan oikeudenhaltijoille.

Edellä esitetyt hyötyskenaariot tulevat konkretisoitumaan ajan myötä, eikä niihin ole tämän vuoksi syytä sukeltaa syvemmälle. Skenaarioiden tärkein anti tutkielmalle on kuitenkin se, että on olemassa perusteita myös sille, että jotkut oikeudenhaltijat päättävät jättää hyödyntämättä DSM-direktiivin 4 artiklan 3 kohdan mukaisen opt out -mekanismin. Lopulta juuri mekanismin käyttöaste oikeudenhaltijoiden keskuudessa määrittää 4 artiklan onnistuneisuuden yksityisen sektorin innovoinnin kannustamisessa. Mikäli mekanismin hyödyntäminen muodostuu pääsääntöiseksi toimintatavaksi läpi toimialojen, koko 4 artiklan merkitys saattaa edellä viitattua Rosatin näkemystä heijastellen vesittyä. Tällöin jäädään varsin kauas oikeudenhaltijoiden ja heidän aineistoaan kaupallisessa tarkoituksessa tekstin- ja tiedonlouhintaan käyttävien tahojen etunäkökohtien välisestä tasapainosta. Mikäli jako mekanismia käyttäviin ja sen käyttämättä jättäviin oikeudenhaltijoihin on kuitenkin edes jossain määrin tasainen tai jopa toiseen suuntaan kallellaan voidaan nähdäkseni katsoa, että

artikla on tuonut parannuksen aiempaan oikeustilaan nähden oikeusvarmuuden ohella myös kannustavuudessa yksityisen sektorin innovointiin.

4. Tekstin- ja tiedonlouhinnan sääntely Yhdysvalloissa

4.1 Relevantin tekijänoikeussääntelyn lähtökohdista

Yhdysvaltojen tekijänoikeussääntelyä on pidetty eurooppalaista vastinettaan myönteisempänä tekstin- ja tiedonlouhinnan kannalta.¹²³ Erityisesti Yhdysvalloissa hyödynnettävä fair use -säännös mahdollistaa verrattain joustavan lähestymistavan tekijänoikeuden rajoituksiin ja poikkeuksiin.¹²⁴ Yhdysvalloissa tekstin- ja tiedonlouhinnan oikeudelliset raamit voidaankin ainakin tiettyyn pisteeseen saakka määrittää analysoimalla kyseisen säännöksen hyödyntämiseen perustuneita tuomioistuinratkaisuja.¹²⁵ Säännöksen ilmaisemasta periaatteesta käytetään myöhemmin rinnasteisesti myös suomenkielistä nimitystä ”kohtuullinen käyttö”.

Kohtuullisen käytön periaatteesta on säädetty Yhdysvaltojen tekijänoikeuslain osiossa 107 seuraavasti:

*”Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include: 1. the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; 2. the nature of the copyrighted work; 3. the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and 4. the effect of the use upon the potential market for or value of the copyrighted work. The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors”.*¹²⁶

Lainattu osio 107 käytännössä kodifioi kohtuullisen käytön doktriinin, joka on kehittynyt Yhdysvalloissa tuomioistuin vetoisesti jo 1800-luvun puolesta välistä alkaen.¹²⁷ Kuten jo

¹²³ Rosati 2019, s. 15.

¹²⁴ Geiger – Izyumenko 2019, s. 6.

¹²⁵ Geiger ym. 2018, s. 10-11.

¹²⁶ Copyright Act of 1976, 17 U.S.C. § 107 (2019).

¹²⁷ Ginsburg 2020, s. 265.

lainauksen ensimmäisestä virkkeestä käy ilmi, fair use -säännös on kaikessa yksinkertaisuudessaan poikkeus Yhdysvaltojen tekijänoikeuslain osioissa 106 ja 106A määritelyihin yksinoikeuksiin. Säännöksen mukaan nimittäin tekijänoikeudella suojatun teoksen kohtuullinen käyttö esimerkiksi valmistamalla alkuperäisestä teoksesta kappaleita tiettyä tarkoitusta varten ei ole tekijänoikeuden loukkaus.

Kohtuulliseen käyttöön liitettyjä hyväksyttyjä tarkoituksia on lueteltu fair use -säännöksessä esimerkinomaisesti. Harkitessaan kohtuullisen käytön toteutumista tuomioistuimen on otettava huomioon säännöksessä mainitut neljä kriteeriä: 1. teoksen käytön tarkoitus ja luonne (the purpose and character of the use); 2. tekijänoikeudella suojatun teoksen luonne (the nature of the copyrighted work); 3. teoksesta käytetyn osan määrä ja oleellisuus suhteessa tekijänoikeudella suojattuun teokseen kokonaisuudessaan (the amount and substantiality of the portion used in relation to the copyrighted work as a whole); 4. teoksen käytön vaikutus tekijänoikeudella suojatun teoksen potentiaaliseen markkinaan tai arvoon (the effect of the use upon the potential market for or value of the copyrighted work). Kriteereiden listaa ei ole kuitenkaan tarkoitettu eksklusiiviseksi taikka yksittäisiä kriteereitä aikaa kestävämmän spesifeiksi.¹²⁸ Näin ollen voitaneen tulkita, että tuomioistuin voi harkintansa mukaan ottaa huomioon myös muita vaatimuksia.

Tuomioistuinten tulee teoksen käytön tarkoitusta ja luonnetta arvioidessaan suhteuttaa arviointi teoksen käytön kaupallisuuteen ja transformatiivisuuteen.¹²⁹ Juuri teoksen käytön transformatiivisuus on noussut viime aikojen oikeuskäytännössä merkittävään asemaan ja asettui siihen myös jäljempänä käsiteltävässä *Google Books* -tapauksessa. Tämän vuoksi on syytä esitellä teoksen transformatiivisen käytön perusteita jo tässä yhteydessä. Muita fair use -säännökseen sisällytettyjä kriteereitä käsitellään sen sijaan yksityiskohtaisemmin *Google Books* -tapausta analysoitaessa. Myös teoksen transformatiivisen käytön konseptia täsmennetään *Google Books* -analyysin yhteydessä oikeuskäytännön avulla.

Levalin mukaan fair use -doktriinin perimmäisen tarkoituksen kannalta teoksen käyttöä voidaan pitää transformatiivisena silloin, kun käyttö antaa alkuperäiselle teokselle lisäarvoa esimerkiksi sitä kautta, että alkuperäistä teosta käytetään raaka-aineena uuden informaation, näkemysten tai havaintojen luomisessa.¹³⁰ Tämän kuvauksen perusteella tekstin- ja

¹²⁸ Sag 2012, s. 54.

¹²⁹ Oracle III, 886 F. 3d, s. 1196.

¹³⁰ Leval 1990, s. 1111.

tiedonlouhinnan voisi katsoa perustuvan melko säännönmukaisesti nimenomaan teosten transformatiiviseen käyttöön. Leval on edelleen katsonut, että tekijänoikeuden alaisen teoksen käyttö voi olla transformatiivista vain silloin, kun käyttö eroaa sekä käyttötavaltaan että -tarkoitukseltaan alkuperäisen teoksen käytöstä.¹³¹ Ginsburg on kuitenkin todennut tuoreen oikeuskäytännön osoittavan, että tuomioistuinten tulkinta teoksen transformatiivisesta käytöstä on laajentunut; sen sijaan, että etsittäisiin näyttöä tukemaan sitä, että alkuperäistä teosta käyttämällä on saatu aikaan transformatiivinen teos, pyritäänkin havaitsemaan transformatiivinen tarkoitus alkuperäisen teoksen käytön taustalla.¹³² Käytännössä tämä tulkinnallinen muutos tarkoittaa sitä, että transformatiivisuuden arvioinnissa kiinnitetään enemmän huomiota alkuperäisen teoksen transformatiiviseen uuteen käyttötarkoitukseen kuin siihen, onko alkuperäisen teoksen pohjalta luotu transformatiivinen uusi teos.¹³³

Alkuperäisen teoksen transformatiivista käyttötarkoitusta painottava suuntaus perustuu Ginsburgin mukaan *Bill Graham Archives v Dorling Kindersley Lt* -tapaukseen.¹³⁴ Tapauksessa tuomioistuin päätyi tulkintaan, jonka mukaan The Grateful Dead -bändin julisteista elämäkerralliseen kahvipöytäkirjaan pienennetyt kokonaiset kuvat olivat transformatiivisia, koska kirja käytti julisteiden kuvia historiallisina artefakteina dokumentoidakseen bändin konsertteja sen sijaan, että kuvia oltaisiin käytetty julisteiden alkuperäisessä, esteettisessä tarkoituksessa.¹³⁵ Juuri tällainen dokumentoinnin ja esteettisyyden tarkoitukset erottava teoksen transformatiivisen käytön arviointi on Ginsburgin mukaan johtanut fair use -doktriinin soveltamisen laajentumiseen kohti uusia teknologioita käyttäviä teoksia siten, että käytön lopputuloksena ei luoda uutta teosta.¹³⁶ Google Books -tapaus on oivallinen tekstin- ja tiedonlouhintaan liittyvä esimerkki tästä kehityksestä.

4.2 Google Books -tapaus: tekstin- ja tiedonlouhinta on kohtuullista käyttöä

4.2.1 Google Books Library Project

Kirjojen digitointiin liittyvät tekijänoikeudelliset haasteet ovat saaneet osakseen merkittävää huomiota viimeisen vuosikymmenen aikana. Tämän taustalla ovat olleet erityisesti

¹³¹ Leval 1990, s. 1111.

¹³² Ginsburg 2020, s. 268.

¹³³ Ginsburg 2020, s. 268.

¹³⁴ Ginsburg 2020, s. 275.

¹³⁵ *Bill Graham Archives v Dorling Kindersley Ltd*, 448 F. 3d 605, s. 609 ja 610.

¹³⁶ Ginsburg 2020, s. 275.

digitointiin tähänneet pienen ja suuren kokoluokan projektit ja niihin liittyneet oikeudelliset epäselvyydet.¹³⁷ *Google Books* -tapaus¹³⁸ sai alkunsa juuri tällaisen projektin tekijänoikeudellisten rajanvetojen selvittämistarpeesta.

Google julkisti vuonna 2004 *Google Books Library Projectin*,¹³⁹ jonka tavoitteena oli skannata ja digitoida suuri määrä tiettyjen maailman johtavien yliopistojen kirjastokokoelmissa olleista teoksista.¹⁴⁰ *Library Project* antoi ja antaa käyttäjille mahdollisuuden suorittaa *Google Books* -sivustolla hakuja miljoonista eri kielillä kirjoitetuista kirjoista.¹⁴¹ Mikäli projektiin sisällytetyn kirjan katsotaan kuuluvan osaksi *public domainia*, Google saattaa sen kokonaisuudessaan yleisön saataville, ja se on näin luettavissa alusta loppuun saakka. Tilanteissa, joissa teos on tekijänoikeussuojan alainen, käyttäjän on mahdollista suorittaa hakuja teoksen tekstistä, mutta *Google Books* näyttää ainoastaan muutamia lyhyitä otteita niistä kohdista tekstiä, joissa haettu termi esiintyy. Tekijänoikeuden haltijan luvalla tästä pääsäännöstä voidaan luonnollisesti poiketa käyttäjämyönteisempään suuntaan.¹⁴²

Google tekee yhteistyötä useiden kirjastojen kanssa ympäri maailman projektin toteuttamiseksi. Yhteistyökirjastot saavat Googelta vastineeksi digitaalisen kopion jokaisesta niiden kokoelmista skannatusta kirjasta, jotta ne voivat puolestaan saattaa digitaaliset kopiot asiakkaidensa saataville tekijänoikeuslain asettamien raamien puitteissa.¹⁴³ Yksinkertaisesti ilmaistuna kyse on sopimusjärjestelystä, johon perustuen Google saa skannattavakseen ja digitoitavakseen sopimukseen sitoutuneiden yliopistojen kirjastojen teosten fyysisiä teoskappaleita, kun taas kirjastot saavat haltuunsa Googlen teoksista laatimat digitaaliset kopiot. Järjestely mahdollistaa sekä *Google Books* -sivuston kehittämisen että kirjastojen digitaalisen teoskannan laajentamisen.

4.2.2 *Google Books* -tapauksen käsittely

Google Books -tapauksen oikeuskäsittely alkoi vuonna 2005, kun yhdysvaltalainen kirjailijoita edustava yhdistys *Authors Guild* ja kustantajia edustava yhdistys *Association of*

¹³⁷ Panezi 2018, s. 597.

¹³⁸ Mainittua nimeä (*Google Books* -tapaus) käytettäessä viitataan tapaukseen kokonaisuudessaan, eikä niinkään sen prosessuaaliseen historiaan kuuluviin eri oikeusasteiden tuomioihin. Näistä tuomioista käytetään lähdeluettelon mukaisesti ilmaisuja *Google Books I* ja *Google Books II*.

¹³⁹ Jatkossa myös nimellä "Library Project".

¹⁴⁰ Lehdistötiedote, Google: Google Checks Out Library Books.

¹⁴¹ *Google Books Library Project*.

¹⁴² *Google Books Library Project*.

¹⁴³ *Google Books Library Project*.

American Publishers haastoivat Googlen oikeuteen.¹⁴⁴ Kanteen perusteena oli kantajien näkemys siitä, että Google Books Library Project loukkasi skannattujen ja digitoitujen teosten tekijänoikeuksien haltijoiden yksinoikeuksia. Väitetyt tekijänoikeusloukkaukset liittyivät Googlen toteuttamaan kirjojen laajamittaiseen digitointiin ilman tekijänoikeuksien haltijoiden lupaa, hakutoiminnolla varustetun tietokannan luomiseen digitoiduista teoksista sekä käyttäjille annettuun mahdollisuuteen lukea lyhyitä otteita näistä teoksista.¹⁴⁵ Googlen vasta-argumentaatio perustui läpi tapauksen käsittelyn fair use -säännöksen varaan.¹⁴⁶ Kustantamot ja Google saavuttivat asiassa sovinnon vuonna 2012. Authors Guild ja Google eivät saaneet sovintoa aikaiseksi, minkä seurauksena yhdistys piti kanteensa voimassa.¹⁴⁷

Sivuuttaen Google Books -tapauksen tutkielman kannalta epäoleellisia prosessuaalishistoriallisia seikkoja voitaneen todeta, että tapauskokonaisuus sisälsi ainakin fair use -arvioinnin kannalta kaksi merkittävää tuomiota. Niistä ensimmäisen, jota kutsuttakoon jatkossa nimellä *Google Books I*, antoi New Yorkin eteläisen alueen piirioikeus. Kyseisessä asiassa olivat kantajina Authors Guildin ohella myös eräät yksityishenkilöt, joilla oli tekijänoikeus joihinkin Googlen digitoimiin kirjoihin.

Tuomioistuin löysi viisi Library Projectin hyötynäkökohtaa, jotka se esitteli ennen varsinaisen fair use -arvioinnin suorittamista. Hyödyiksi katsottiin seuraavat seikat: 1. Google Booksin nähtiin tarjoavan lukijoille ja tutkijoille uuden ja tehokkaan tavan löytää kirjoja; 2. Google Books edistää tekstin- ja tiedonlouhintaan perustuvaa tutkimusta mahdollistamalla muun muassa eri sanojen esiintyvyyden, syntakstisten mallien ja kirjallisuuden tyylin muuttumisen tarkemman analysoinnin; 3. Google Books tasa-arvoistaa ”pääsyä käsiksi” kirjallisuuteen etenkin sellaisten henkilöiden osalta, jotka eivät pysty apulaitetarpeidensa vuoksi lukemaan normaalia painettua tekstiä; 4. Google Books helpottaa kirjojen säilyttämistä ja antaa joillekin niistä ”uuden elämän”; 5. Google Books tuo kirjat laajemmin ihmisten nähtäville ja sitä kautta heidän ostettavakseen.¹⁴⁸

Kantajien kanteet perustuivat säädöstasolla U.S. Copyright Actin 106 §:ään, jossa säädetään muun muassa tekijänoikeuden haltijan yksinoikeudesta valmistaa teoksestaan kappaleita, jakaa kappaleita eteenpäin ja näyttää kappaleita yleisölle.¹⁴⁹ Google puolestaan vetosi

¹⁴⁴ Xalabarder 2014, s. 53.

¹⁴⁵ Matulionyte 2015, s. 47.

¹⁴⁶ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 284.

¹⁴⁷ Xalabarder 2014, s. 53.

¹⁴⁸ Authors Guild, Inc. v. Google Inc., 954 F. Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 287-288.

¹⁴⁹ Copyright Act of 1976, 17 U.S.C. § 106 (2019).

puolustuksenaan aiemmin todetulla tavalla fair use -säännökseen. Tuomioistuimen mukaan tapauksen ainoaksi pääasialliseksi kysymykseksi jäikin näin ollen se, voitiinko Googlen suorittama tekijänoikeudella suojattujen teosten käyttö katsoa fair use -säännöksen nojalla hyväksytyksi. Tuomioistuin päätyi lopulta vastaamaan kysymykseen myöntävästi.¹⁵⁰

Tuomioistuin analysoi fair use -säännöstä kattavasti käyden kaikki säännöksen neljä kriteeriä läpi varsin yksityiskohtaisesti. Ensimmäinen kriteeri edellyttää arviointia tekijänoikeudella suojatun teoksen käytön tarkoituksesta ja luonteesta. Kriteeriä analysoitaessa on edellä kuvatulla tavalla kiinnitettävä huomiota sekä siihen, voidaanko teoksen käyttöä pitää riittävän transformatiivisena että siihen, perustuuko käyttö enemmän kaupalliseen vai ei-kaupalliseen tarkoitukseen.¹⁵¹ Tuomioistuin katsoi, että Googlen toteuttamaa tekijänoikeudella suojattujen teosten käyttöä voitiin pitää erittäin transformatiivisena sekä Google Books -sivuston hakutoiminnon että sivuston näyttämien lyhyiden otteiden osalta. Tekstin- ja tiedonlouhinnan kannalta olennaisin perustelu Google Booksin sisältämien teosten transformatiiviselle käytölle oli se, että Google oli muuttanut kirjojen tekstin dataksi määrällisen tutkimuksen, kuten juuri tekstin- ja tiedonlouhinnan mahdollistamiseksi uusilla tutkimusalueilla. Näin ollen kirjojen sanoja katsottiin käytetyn tavalla, jolla niitä ei ollut ennen käytetty.¹⁵²

Mitä tulee punnintaan teosten kaupallisen tai ei-kaupallisen käyttötarkoituksen vaikutuksesta fair use -säännöksen mukaiseen arviointiin, tuomioistuin otti huomioon, että kaupallisella käyttötarkoituksella oli taipumus johtaa säännöksen soveltamista vastustavaan löydökseen.¹⁵³ Mikäli tällainen lähestymistapa olisi tulkittu kategoriseksi, Googlen kohtuulliseen käyttöön perustunut puolustus ei olisi pitänyt. On nimittäin selvää, että Google on mitä suurimmassa määrin kaupallinen toimija, vaikka sen toteuttamaa teosten käyttöä ei voinutkaan kuvailla välittömästi kaupalliseksi.¹⁵⁴ Tuomioistuin päätyi kuitenkin harkinnassaan tulkintaan, jonka mukaan ensimmäisen kriteerin arviointi puolsi sekä käytön

¹⁵⁰ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 289.

¹⁵¹ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 291.

¹⁵² Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 291-292.

¹⁵³ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 291. Viitattuun virkkeeseen liittyen ks. esim. Harper & Row v. Nation Enterprises, 471 U.S. 539 (1985), (*Harper & Row*) tai Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994), (*Campbell*).

¹⁵⁴ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 291.

transformatiivisuuden että yleisesti käyttötarkoituksen osalta kohtuullisen käytön doktriinin soveltumista tapauksessa.¹⁵⁵

Fair use -säännöksen toinen kriteeri liittyy tekijänoikeudella suojattujen teosten luonteeseen. Google Books -tapauksessa teokset olivat kirjoja. Tuomioistuimien otti ensinnäkin huomioon, että suurin osa Google Booksin käyttöön digitoiduista kirjoista oli luonteeltaan tietokirjallisuutta.¹⁵⁶ Tällä oli arvioinnissa jonkinlaista relevanssia siitä syystä, että kaunokirjallisuuden teoksille oli suotu tietokirjoja korkeampi tekijänoikeudellinen suoja.¹⁵⁷ Toiseksi se totesi, että kirjat oli julkaistu ja ne oli saatettu yleisön saataville.¹⁵⁸ Näillä perusteilla tuomioistuimien katsoi Googlen läpäisevän myös fair use -säännöksen toisen kriteerin kriittisen tarkastelun.¹⁵⁹

Kolmas kriteeri pureutuu siihen, kuinka suurta ja olennaista osaa tekijänoikeudella suojatusta teoksesta on käytetty. Library Project perustui aiemmin kuvatulla tavalla siihen, että kirjastoilta saadut teokset skannattiin ja digitoitui kokonaisuudessaan. Ilman tällaista täysimääräistä kappaleen valmistamisprosessia erityisesti Google Booksin hakutoiminto olisi jäänyt varsin vaillinaiseksi. Tuomioistuimet ovat eräissä tapauksissa kuitenkin linjanneet, että kappaleen valmistaminen kopioimalla tekijänoikeudella suojattu teos kokonaisuudessaan saattaa olla tietyissä tilanteissa välttämätöntä ja siten jopa linjassa kohtuullisen käytön kanssa.¹⁶⁰ Vaikka tuomioistuimien huomioi sekä kyseisen linjauksen että Googlen pyrkimykset rajoittaa hakutuloksena ilmaantuvan tekstin määrää, se päättyi tulkintaan, jonka mukaan Googlen toiminta ei kolmannen kriteerin osalta täysin vastannut kohtuulliselle käytölle asetettuja raameja.¹⁶¹

Neljäs kohtuullista käyttöä määrittävä kriteeri koskee tekijänoikeudella suojatun teoksen käytön vaikutuksia alkuperäisen teoksen markkinatilanteeseen tai -arvoon. Kantajien mukaan Google Books -sivustolla oli negatiivinen vaikutus sinne digitoitujen kirjojen markkinatilanteeseen, ja he katsoivat myös, että Googlen skannatut kappaleet teoksista toimivat alkuperäiset teokset potentiaalisesti korvaavina vaihtoehtoina markkinoilla.¹⁶²

¹⁵⁵ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 292.

¹⁵⁶ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 292.

¹⁵⁷ Stewart v. Abend, 495 U.S. 207 (1990), (*Stewart v. Abend*), s. 237.

¹⁵⁸ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 292.

¹⁵⁹ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 292.

¹⁶⁰ Ks. esim. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984), (*Sony*), s. 449-450.

¹⁶¹ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 292.

¹⁶² Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 292.

Tuomioistuim ei kuitenkaan pitänyt väitteitä perusteltuina, vaan totesi, että Google Books kohensi kirjojen myyntilukuja ja sitä myöten hyödytti ennen kaikkea tekijänoikeuksien haltijoita. Tuomioistuim perusteli kantaansa ensinnäkin sillä, että Google Books auttoi lukijoita ja tutkijoita löytämään teokset helpommin. Toiseksi tuomioistuim lausui, että Google Booksin tarjoamat linkit kirjojen ostamisen mahdollistamiseksi tekivät kirjojen tilaamisesta entistä helpompaa. Näillä perusteilla tuomioistuim totesi neljännen kriteerin analyysin puoltavan vahvasti kohtuullisen käytön havaintoa tapauksessa.¹⁶³

Tuomioistuim totesi Google Books I -tuomion loppupäätelmänä, että Googlen suorittamat toimenpiteet, joita kantajat väittivät tekijänoikeuksiensa loukkauksiksi, olivat kohtuullista käyttöä ja siten hyväksyttyjä poikkeuksia Yhdysvaltojen tekijänoikeuslain osioiden 106 ja 106A suomiin yksinoikeuksiin. Päätelmän perustana oli edellä esitelty neljän kriteerin punninta. Näin ollen kanne Googlen väitetyistä tekijänoikeusloukkauksista päädyttiin hylkäämään.¹⁶⁴ Kantajat valittivat tuomiosta, mutta toisen piirikunnan vetoomustuomioistuim seurasi omassa tuomiossaan pitkälti piirioikeuden antamia perusteluita ja päätöstä.¹⁶⁵ Seuraavaksi käydään läpi sellaisia vetoomustuomioistuimen päätökselleen antamia perusteluita, jotka täsmentävät edellä esitettyä piirioikeuden analyysia. Tämän jälkeen pohditaan lyhyesti tuomioista välittyvää kuvaa yhdysvaltalaisesta tekstin- ja tiedonlouhintaan kohdistuvasta sääntelymallista.

Toisen piirikunnan vetoomustuomioistuim lähestyi Google Books -tapausta¹⁶⁶ piirioikeuden tavoin fair use -säännöksen neljän kriteerin kautta. Vetoomustuomioistuim jakoi ensimmäisen kriteerin käsittelyn kahteen osa-alueeseen, joista ensimmäiseksi se kävi läpi Google Books -sivuston hakufunktion, minkä jälkeen se siirtyi peilaamaan kriteeriä sivuston käyttäjille tarjoamaan mahdollisuuteen nähdä lyhyitä otteita teoksista haettujen sanojen ympäriltä. Hakufunktion osalta vetoomustuomioistuim totesi, että Googlen toimintamalli sisälsi erittäin transformatiivisen tarkoituksperän. Lisäksi tuomioistuim viittasi *HathiTrust*-tapauksessa¹⁶⁷ esittämäänsä näkökulmaan, jonka mukaan sekä digitaalisten kappaleiden

¹⁶³ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 293.

¹⁶⁴ Authors Guild, Inc. v. Google Inc., 954 F.Supp.2d 282 (S.D.N.Y. 2013), (*Google Books I*), s. 294.

¹⁶⁵ Matulionyte 2015, s. 48.

¹⁶⁶ Toisen piirikunnan vetoomustuomioistuimen antamaan tuomioon viitataan leipätekstissä nimellä *Google Books II*, erotuksena aiemmin esittelystä piirioikeuden tuomiosta, josta käytettiin nimeä *Google Books I*.

¹⁶⁷ Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2nd Cir. 2014). Tapaus linkittyi sekä ajallisesti että asiallisesti *Google Books* -tapaukseen. Siinä kantajat haastoivat HathiTrust-nimisen toimijan oikeuteen väitettyjen tekijänoikeusloukkausten johdosta. HathiTrust koostui Library Projectiin osallistuneista kirjastoista, ja sen tarkoitus oli koota yhteen kirjastojen Googelta saamat digitaaliset kappaleet fyysisistä teoskappaleistaan. Myös HathiTrust tarjosi asiakkailleen mahdollisuuden suorittaa hakuja teosten teksteistä

valmistaminen että niiden käyttäminen hakufunktion mahdollistamiseksi voitiin katsoa kohtuulliseksi käytöksi.¹⁶⁸ Tuomioistuin esitti tulkinnalleen teosten käytön transformatiivisuudesta muun muassa seuraavat perusteet: 1. käyttäjän tekemän sanahaun tulos on monella tapaa erilainen kuin se sivu tai kirja, josta haku on tehty; 2. Googlen tarkoitus on antaa käyttäjille tärkeää tietoa kirjoista hakufunktion avulla; 3. Google antaa käyttäjilleen mahdollisuuden saada tietoonsa esimerkiksi tiettyjen sanojen esiintyvyyteen liittyvää dataa.¹⁶⁹ Mitä tulee Googlen Booksin käyttäjilleen tarjoamaan mahdollisuuteen lukea lyhyitä otteita teoksista, vetoomustuomioistuin totesi, että kyseinen toiminnallisuus täsmentää hakufunktion käyttäjälle antamaa hyötyä kuitenkin vaarantamatta liikaa oikeudenhaltijoiden etua. Tämän katsottiin luonnollisesti puoltavan kohtuullista käyttöä painottavaa tulkintaratkaisua.¹⁷⁰

Fair use -säännöksen ensimmäiseen arvioitavaan kriteeriin sisältyy aiemmin kerrotulla tavalla myös tarve analysoida tekijänoikeudella suojatun teoksen käyttöä siltä pohjalta, onko käytön taustalla kaupallinen vai ei-kaupallinen tarkoitus. Yhdysvaltojen korkeimman oikeuden näkemys tähän punnintaan tiivistyi aikanaan *Sony*-tapausta käsitellessä, kun tuomioistuin totesi, että kaikenlainen tekijänoikeudella suojatun materiaalin kaupallinen käyttö on lähtökohtaisesti tulkittava kohtuuttomaksi.¹⁷¹ Vetoomustuomioistuin kuitenkin totesi myös muuhun *Sony*-tapauksen jälkeiseen oikeuskäytäntöön vedoten, että korkeimman oikeuden tapauksessa tekemä tulkinta oli suuresti liioiteltu.¹⁷² Tuomioistuin otti kuitenkin huomioon *Campbell*-tapauksen tulkinnan, jonka mukaan tekijänoikeudella suojatun teoksen käytön taustalla oleva kaupallinen tarkoitus voi johtaa kohtuuttomaan käyttöön erityisesti silloin, kun käytön riittävästä transformatiivisuudesta ei ole onnistuttu esittämään näyttöä.¹⁷³ Tuomioistuin päätyi lopulta kokonaisharkinnassaan siihen, että tässä tapauksessa ei ollut mitään syytä tulkita Googlen yhtiönlaajuisen voitontavoittelutarkoituksen johtavan kohtuuttoman käytön toteamiseen. Etenkään, kun otetaan huomioon Googlen suorittaman

kokonaisuudessaan. Tapausta ei kuitenkaan käsitellä tutkielmassa syvällisemmin, sillä kyseessä ei ole kaupallisen toimijan toteuttama projekti.

¹⁶⁸ Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 217.

¹⁶⁹ Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 217.

¹⁷⁰ Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 217-218.

¹⁷¹ Sony Corporation of America v. Universal City Studios, Inc., 464 U.S. 417 (1984), s. 451: "Every commercial use of copyrighted material is presumptively... unfair."

¹⁷² Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 218. Ks. viitatussa virkkeessä mainitusta *Sony*-tapauksen jälkeisestä oikeuskäytännöstä esim. *Campbell*, s. 583-584.

¹⁷³ *Campbell*, s. 579; Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 219.

teosten käytön transformatiivisuus ja se, ettei Google Booksin voida katsoa kilpailevan samoilla markkinoilla teosten tekijänoikeuksien haltijoiden kanssa.¹⁷⁴

Toisen kriteerin osalta vetoomustuomioistuin tyytyi toteamaan, että tekijänoikeudella suojattujen teosten luonteiden arviointi ei tässä tapauksessa juurikaan vaikuttanut kohtuullisen ja kohtuuttoman käytön väliseen kokonaisarvioon. Näin oli siitä huolimatta, että kirjat olivat luonteeltaan tietokirjoja, joiden tekijänoikeussuojaa voitiin piirioikeuden tuomiossa kuvatulla tavalla pitää matalampana kuin kaunokirjallisilla teoksilla.¹⁷⁵

Vetoomustuomioistuin jakoi myös kolmannen kriteerin käsittelyn kahteen osaan, joista ensimmäisessä se analysoi Google Booksin hakufunktiota ja toisessa sivuston käyttäjilleen tarjoamaa mahdollisuutta nähdä lyhyitä otteita teoksista. Hakufunktion osalta tuomioistuin totesi, että tuomioistuimet eivät ole yleisesti omaksuneet sellaista kategorista tulkintaa, jonka mukaan tekijänoikeudella suojatun teoksen kopioiminen kokonaisuudessaan johtaisi automaattisesti havaintoon kohtuuttomasta käytöstä.¹⁷⁶ Sen sijaan tuomioistuin katsoi tapausta edeltäneeseen oikeuskäytäntöön vedoten, että kopioinnin laajuutta ja olennaisuutta arvioitaessa on kiinnitettävä huomiota myös tekijänoikeudella suojattujen teosten käytön transformatiivisuuteen ja siihen, kuinka helposti käytettyjen teosten kopiot voidaan katsoa alkuperäiset teokset korvaaviksi vaihtoehdoiksi markkinoilla.¹⁷⁷ Näitä tekijöitä arvioiden tuomioistuin päätyi tulkintaan, jonka mukaan Google läpäisi fair use -arvioinnin kolmannen kriteerin hakufunktionsa osalta.

Lyhyiden otteiden tarjoamisen osalta tuomioistuin totesi ensinnäkin, että on tärkeämpää tarkastella tekijänoikeudella suojatun teoksen käytön laajuutta ja olennaisuutta suhteessa siihen, kuinka suuri osa kopioidusta materiaalista päädytään saattamaan yleisön saataville kuin sitä, kuinka suurta ja/tai olennaista osaa tekijänoikeuden alaisesta teoksesta on kopioitu. Tällöin saadaan helpommin selville olennaisin tieto, eli se, kuinka korvaavaksi kopioitu teos voidaan alkuperäiseen teokseen nähden katsoa relevanteilla markkinoilla.¹⁷⁸ Tuomioistuin lausui, että Google oli rakentanut lyhyitä otteita käyttäjilleen tarjoavan toiminnallisuuden tavalla, joka pyrki erilaisten teknisten metodien avulla nimenomaan välttämään tilannetta, jossa Googlen skannaama ja digitoima kopio voitaisiin katsoa alkuperäisen teoksen kanssa

¹⁷⁴ Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 219.

¹⁷⁵ Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 220.

¹⁷⁶ Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 221.

¹⁷⁷ Campbell, s. 586-587; Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 221-222.

¹⁷⁸ Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 222.

kilpailevaksi vaihtoehdoksi potentiaalisilla markkinoilla.¹⁷⁹ Näin ollen myös lyhyitä otteita käyttäjille tarjoava Google Booksin toiminnallisuus läpäisi fair use -säännöksen kolmannen kriteerin tulkinnalliset reunaehdot.

Harper & Row -tapauksessa Yhdysvaltojen korkein oikeus kuvasi fair use -säännöksen neljättä kriteeriä koko kohtuullisen käytön doktriinin tärkeimmäksi elementiksi.¹⁸⁰ *Campbell*-tapauksessa sama oikeusaste puolestaan painotti neljännen kriteerin läheistä suhdetta ensimmäisen kriteerin kanssa; mitä selvemmin kopiointi suoritetaan tarkoituksessa, joka eroaa alkuperäisen teoksen tarkoituksesta, sitä epätodennäköisempää on, että kopio voidaan katsoa kykeneväksi korvaamaan alkuperäisen teoksen markkinoilla.¹⁸¹ Google Books II -tapauksessa vetoomustuomioistuimien kuitenkin muistutti, että vaikka teoksen kopioinnin tarkoitusta voitaisiinkin pitää transformatiivisena, kopiointi saattaa silti laskea alkuperäisen tekijänoikeudella suojatun teoksen arvoa, mikäli se toteutetaan tavalla, joka johtaa laajamittaiseen alkuperäisen teoksen merkittävien osien paljastamiseen.¹⁸² Tältä osin vetoomustuomioistuimen arvioinnin kohteena oli hakufunktion sijaan Google Booksin käyttäjilleen tarjoama mahdollisuus nähdä lyhyitä otteita teoksista.

Tuomioistuimien päätyi arvioinnissaan siihen, että mainittu sivuston toiminnallisuus ei muodostanut korvaavaa vaihtoehtoa alkuperäiselle teokselle taikka vähentänyt merkittävästi alkuperäisen teoksen arvoa siitä huolimatta, että sen voidaankin arvella joissakin tapauksissa aiheuttavan pientä vähentymistä alkuperäisen teoksen myyntiluvuissa. Lisäksi tuomioistuimien katsoi, että myynnin vähentymistä koskevia tilanteita voitiin yleisesti ottaen olettaa tapahtuviksi silloin, kun tutkijan pyrkimyksenä oli löytää tekstistä ainoastaan jokin historiallinen fakta. Niitä ei puolestaan ole mahdollista suojata tekijänoikeudella, minkä vuoksi seikalla ei ollut tuomioistuimen harkintaa liikauttavaa merkitystä.¹⁸³ Näillä perusteilla vetoomustuomioistuimien katsoi myös neljännen kriteerin läpäistyksi ja totesi Googlen toiminnan kuuluneen kokonaisuudessaan kohtuullisen käytön piiriin.¹⁸⁴

Piirioikeus ja vetoomustuomioistuimien päätyivät siis samoihin lopputuloksiin Google Books -tapauksessa. Näin oli siitäkin huolimatta, että niiden tulkinnat kohtuullisen käytön toteutumisesta erosivat hienoisesti toisistaan kolmannen kriteerin kohdalla. Toisaalta oli

¹⁷⁹ Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 222.

¹⁸⁰ Harper & Row, s. 566.

¹⁸¹ Campbell, s. 591.

¹⁸² Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 223.

¹⁸³ Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 224.

¹⁸⁴ Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir. 2015), s. 224-225.

helposti havaittavissa, että tuomioistuimet olivat keskenään samaa mieltä ensimmäisen ja neljännen kriteerin tulkinnasta, joita voitaneen pitää tapauksen ratkaisun kulmakivinä.

4.3 Fair use -kriteerien arviointia kaupallisen tekstin- ja tiedonlouhinnan näkökulmasta

Google Books -tapauksen ratkaisu tarkoittaa nähdäkseni sitä, että tekstin- ja tiedonlouhinnan kaupallinen tarkoitus ei suoraan johda siihen, että alkuperäisen teoksen käyttö todettaisiin luonteeltaan kohtuuttomaksi. Sen sijaan onkin niin, että mikäli teoksen käytön tarkoitus ja luonne todetaan riittävän transformatiivisiksi, voidaan käyttö katsoa kohtuulliseksi siitä huolimatta, että käytön taustalla ovat vaikuttaneet kaupalliset tarkoitukset. Ginsburg onkin todennut Yhdysvaltojen korkeimman oikeuden Campbell-tapauksessa esittämien perustelujen osoittavan, että mitä transformatiivisemmasta teoksen käytöstä on kyse, sitä pienempi merkitys on teoksen käytön kaupallisella tarkoituksella.¹⁸⁵

Google Books -sivuston Ngram Viewer¹⁸⁶ on tekstinlouhintaan perustuva hakutyökalu, jolle käyttäjä voi syöttää fraaseja tarkoituksessa saada tieto kyseisten fraasien esiintyvyydestä kirjoissa, jotka Google on kerännyt digitaaliseen tietokantaansa Library Projectin ja Partner Programmen¹⁸⁷ kautta. Ngram Viewer mahdollistaa näin ollen myös edellä tuomioistuinratkaisujen esittelyn yhteydessä mainittujen syntaktisten mallien ja kirjallisuuden tyylien kehittymisen tutkimisen. Osana Google Books -tapauksista myös Ngram Viewer luonnollisesti todettiin tekijänoikeudella suojattujen teosten kohtuulliseksi käytöksi.

Nähdäkseni ratkaisu seuraa samaa linjaa edellä esiteltujen Levalin näkemysten kanssa. Hakujen ja tekstinlouhinnan suorittamisen mahdollistaminen digitoiduista kirjoista voitaneen nimittäin katsoa selkeäksi esimerkiksi tilanteesta, jossa alkuperäistä teosta käytetään pohja-aineena uuden tiedon tuottamisessa. Tämän lisäksi teoksen käyttö eroaa tuomioistuintenkin toteamalla tavalla alkuperäisten teosten käyttötavasta ja -tarkoituksesta. Myös *Sag* on katsonut, että Google Books -tapaukseen liittyvät päätökset toivat esiin myönteisen suhtautumisen tekstin- ja tiedonlouhintametodien käyttämiseen luotaessa uusia

¹⁸⁵ Ginsburg 2020, s. 268.

¹⁸⁶ Ks. lisää esim. Google Books Ngram Viewer.

¹⁸⁷ Google Books Partner Programme on Google Books -sivuston toinen osa Library Projectin ohella. Partner Programme tarjoaa kustantajille ja kirjailijoille ilmaisen mahdollisuuden mainostaa kirjojaan verkossa Google Books -sivuston kautta. Tietyissä maissa osaksi Partner Programmea kuuluu myös mahdollisuus myydä kirjojen digitaalisia versioita Google Play -palvelussa, jolloin myyjä saa luonnollisesti myös tuloja teoksensa myynnin mukaisesti. Ks. lisää esim. About the Google Books Partner Program.

näkemyksiä ja havaintoja alkuperäisistä teoksista.¹⁸⁸ Nähdäkseni päätökset sinetöivät lähtökohdan, jonka mukaan tekstin- ja tiedonlouhinnan tarpeita vastaava alkuperäisen teoksen käyttö voidaan tulkita transformatiiviseksi ja näin ollen siltä osin fair use -säännöksen ensimmäisen kriteerin läpäiseväksi. Myös Sag on tukenut tällaista näkemystä katsoen, että ei-ekspressiiviset alkuperäisten teosten käyttötavat eivät uhkaa alkuperäisiä teoksia korvattavuudella, sillä ne ainoastaan tuottavat tietoa alkuperäisestä teoksesta eivätkä viesti sen sisältämää tarkoitusta.¹⁸⁹

Tutkielman kannalta ensimmäisen kriteerin kaupallista tarkoitusta tekijänoikeudella suojatun teoksen käytössä koskevan osion tulkitseminen on erityisen kiinnostavaa. Google Books -tapauksen tuomiot ja erityisesti niiden perustelut osoittavat asian nykyhetken tulkinnasta ainakin sen, että kaupallinen tarkoitus ei kategorisesti sulje ovea kohtuullisen käytön toteamiselta. Sen sijaan tuomioistuimet vaikuttivat antavan Googlen asemalle kaupallisena toimijana vähemmän negatiivista painoarvoa sillä perusteella, että kyseessä ollut alkuperäisten teosten käyttö todettiin erittäin transformatiiviseksi. Tämä ei itsessään ole kuitenkaan poikkeuksellista, sillä tuomioistuinten on usein todettu painottavan kaupallisen tarkoituksen aspektia sitä vähemmän, mitä transformatiivisemmasta teoksen käytöstä on ollut kyse.¹⁹⁰ Sag on jopa katsonut, että koska tekstinlouhinta ja muut ei-ekspressiiviset tekijänoikeuden alaisen teoksen käyttötavat eivät voi korvata alkuperäisen teoksen tekijän ilmaisua, käyttötavan kaupallisuutta ei voida pitää kovinkaan informatiivisena pohdittavana kriteerinä.¹⁹¹ Lienee siis tulkittavissa, että kaupallisessa tarkoituksessa toteutettava tekstin- ja tiedonlouhinta ei ole kategorisesti kiellettyä Yhdysvaltojen tekijänoikeudellisessa käytännössä. Oikeastaan tilanne vaikuttaa päinvastaiselta: luonteeltaan mitä ilmeisimmin transformatiiviseksi luokiteltava alkuperäisen teoksen käyttö tekstin- ja tiedonlouhintaa varten johtaa tilanteeseen, jossa toiminnan kaupallisuuden arvioinnille annetaan melko vähän painoarvoa.

Fair use -säännöksen toinen, tekijänoikeuden alaisen teoksen luonnetta koskeva kriteeri ei ole ainakaan Google Books -tapauksen valossa yhtä merkittävä tekstin- ja tiedonlouhinnan sallittavuuden kannalta kuin esimerkiksi ensimmäinen kriteeri. Molemmissa edellä esitellyissä tuomioissa keskityttiin analysoimaan, millainen merkitys kirjojen kuulumisella

¹⁸⁸ Sag 2019, s. 319.

¹⁸⁹ Sag 2019, s. 320.

¹⁹⁰ Diaz 2013, s. 692.

¹⁹¹ Sag 2019, s. 323.

tietokirjallisuuden luokkaan oli kohtuullisen käytön arvioinnin kannalta. Tältä osin verrokiksi otettiin luonnollisesti kaunokirjallisten teosten luokka. Sagin mukaan tekijänoikeudella suojatun teoksen luonne on tärkeä analysoitava seikka kohtuullisen käytön arvioinnissa, mutta tekstinlouhinnan laillisuutta ei tästä huolimatta määritä yksinään se, ovatko louhinnan kohteena olevat teokset luonteeltaan luovia vai informatiivisia.¹⁹²

Kuten todettua, tekijänoikeudella suojatun teoksen käytön laajuus ja olennaisuus ovat olennaisia mittareita fair use -analyysin kannalta. Näin lienee myös tekstin- ja tiedonlouhinnan osalta, vaikka onkin niin, että louhinta edellyttää suurimmassa osassa tapauksista välttämättömästi teoksen kopioimista kokonaisuudessaan. Tämän totesi myös vetoomustuomioistuimien Google Books I -tapauksessa antamassaan tuomiossa, jossa tekijänoikeuden alaisten teosten kokonaisvaltainen kopiointi katsottiin kohtuulliseksi käytöksi, kun otettiin huomioon, että se oli välttämätöntä, jotta käytön transformatiivinen tarkoitus saatiin täytettyä. Fair use -säännöksen kolmannen kriteerin voidaankin katsoa mittaavan yhtä lailla laadullisia kuin määrällisiä parametreja.¹⁹³ Mitä tulee tekstin- ja tiedonlouhinnan kaltaisiin ei-ekspressiivisiin käyttötapoihin, on argumentoitu, että teoksen kopioiminen kokonaisuudessaan voidaan katsoa laadullisen analyysin kannalta hyväksyttäväksi.¹⁹⁴ Tulkinta viestii nähdäkseni joustavasta suhtautumisesta tekijänoikeudella suojatun teoksen käytön laajuuden ja olennaisuuden väliseen rajanvetoon. Lieneekin niin, että erilaiset teosten käyttötavat johtavat erilaisiin valintoihin siitä, painotetaanko kolmannen kriteerin arvioinnissa enemmän teoksen käytön määrää ja laajuutta vai käytön olennaisuutta. Tekstin- ja tiedonlouhinnan osalta on luonnollista, että arviointi kallistuu käytön olennaisuuteen, sillä määrällisesti teoksen käyttö kattaa usein teoksen kokonaisuudessaan.

Myöskään neljäs kriteeri ei heikennä tekstin- ja tiedonlouhinnan sallittavuutta kohtuullisen käytön näkökulmasta. Eräänlaiseen markkina-analyysiin perustuvan kriteerin avulla pyritään määrittämään, onko tekijänoikeuden alaisen teoksen käytössä luotu samoilla markkinoilla alkuperäisen teoksen kanssa kilpaileva vaihtoehto tai onko alkuperäisen teoksen arvo muutoin heikentynyt sen käytön kautta. Piirioikeuden ja vetoomustuomioistuimen mukaan kumpikaan skenaarioista ei toteutunut Google Books -tapauksessa. Kuten muidenkin kriteerien osalta, transformatiivisuus alkuperäisen teoksen

¹⁹² Sag 2019, s. 324-325.

¹⁹³ Diaz, s. 694.

¹⁹⁴ Sag, s. 326.

käytössä antaa myös neljännen kriteerin kohdalla signaalin kohtuullisen käytön löydöksestä. Sag onkin todennut, että hakukoneen tietystä verkkosivusta tai kirjaston kirjasta tuottama informaatio ei voi mitenkään muodostaa korvaavaa vaihtoehtoa verkkosivun tai kirjan tekstin sisältämälle informaatiolle saatikka alkuperäiselle, tekijänoikeudella suojatulle ilmaisulle kyseisissä teksteissä. Hän on katsonut saman pätevän siinäkin tapauksessa, että tekijänoikeuden haltija tuottaisi tekstin- ja tiedonlouhintaan keskittyviä palveluita markkinoille.¹⁹⁵

Kokonaisuudessaan Google Books -tapauksen lopputuloksesta huokuu teosten käytön transformatiivisuutta painottava tulkinta jopa siinä määrin, että muut kohtuulliselle käytölle annetut kriteerit tuntuvat hukkuvan sen alle. Myös kirjallisuudessa on katsottu teosten transformatiiviseen käyttöön perustuvan analyysin muuttaneen fair use -säännöksen mukaisen neljän kriteerin testin ennemminkin yhden kriteerin arvioinniksi.¹⁹⁶

4.4 Vertailua EU:n oikeustilaan nähden

Tekstin- ja tiedonlouhinnan tekijänoikeudelliset raamit on asetettu eri tavoin Yhdysvalloissa ja EU:ssa. Järjestelmien eroavaisuudet eivät nähdäkseni perustu uuteen DSM-direktiiviin, vaan juontavat ennemmin juurensa Yhdysvalloissa käsitellystä Google Books -tapauksesta ja sen ratkaisemisesta hyödynnetystä fair use -säännöksestä. Toisaalta on myös kiistatonta, että EU:n tekijänoikeusjärjestelmän tekstin- ja tiedonlouhintaa koskevat elementit ovat kokeneet suuren muutoksen DSM-direktiivin myötä. Aiemmin kovin hajanainen ja oikeudellisesti epäselvä tietoyhteiskuntadirektiiviin, tietokantadirektiiviin ja tietokoneohjelmadirektiiviin perustunut säädöspohja korvautui yksinomaan tekstin- ja tiedonlouhintaa koskevilla DSM-direktiivin 3 ja 4 artikloilla.

Bentham on katsonut, että oikeusjärjestelmässä on olemassa sekä tarkkoja että löyhiä ja yleisiä sääntöjä.¹⁹⁷ Jälkimmäisiä sääntöjä voidaan ainakin tässä yhteydessä kutsua myös standardeiksi. *Kaplow* on määritellyt sääntöjen ja standardien väliseksi eroksi sen, että siinä missä yksittäinen sääntö määrittelee tarkasti, millainen käyttäytyminen on sallittua ja millainen kiellettyä, standardi jättää vastuun sallitun ja ei-sallitun käyttäytymisen tarkemmasta määrittelystä tuomioistuimelle.¹⁹⁸ *Hughes* on todennut, että Yhdysvaltojen fair use -säännös on oiva esimerkki standardista ja lisännyt, että rationaalisen lainsäätäjän

¹⁹⁵ Sag, s. 327.

¹⁹⁶ Ginsburg 2020, s. 268.

¹⁹⁷ Bentham 1891, s. 158.

¹⁹⁸ Kaplow 1992, s. 560.

näkökulmasta tekijänoikeusjärjestelmä on enemmän kuin sopiva ympäristö standardien käyttämiselle sääntöjen sijaan.¹⁹⁹ Kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan näkökulmasta fair use -säännös vaikuttaakin olevan luonteeltaan ennemmin standardi kuin sääntö. On nimittäin niin, että säännöksessä on ainoastaan asetettu raamit sille, millä perusteilla tekijänoikeudella suojattua aineistoa on mahdollista käyttää ilman oikeudenhaltijan suostumusta.

DSM-direktiivin 4 artiklaa voitaneen luonnehtia ennemmin säännöksi kuin standardiksi. Nähdäkseni artikkelissa on asetettu melko tarkat rajat sille, millä edellytyksillä tekstin- ja tiedonlouhintaa saa toteuttaa tekijänoikeudella suojatusta aineistosta muussa kuin ei-kaupallisessa tarkoituksessa. Toisaalta myös DSM-direktiivin 4 artiklaan sisältyy sellaisia jossain määrin avoimia ilmaisuja, joita voidaan olettaa täsmennettävän jatkossa esimerkiksi tuomioistuinratkaisujen kautta. Tällaisia ilmaisuja ovat tai sellaisiksi voivat muodostua artiklan 1 kohdan termi ”laillinen pääsy”, 2 kohdassa teosta hyödyntävälle taholle annettu mahdollisuus säilyttää valmistettuja kappaleita ja kopioita niin kauan, kuin tekstin- ja tiedonlouhintaa varten on tarpeen sekä 3 kohdassa mainittu oikeuksien pidättäminen ”asianmukaisella tavalla”. On kuitenkin niin, että siinä missä DSM-direktiivin 4 artikla sääntelee ainoastaan kaupallisessa tarkoituksessa toteutettavaa tekstin- ja tiedonlouhintaa, Yhdysvaltojen fair use -säännöksellä luodaan mahdollisuus oikeudenhaltijan yksinoikeudesta poikkeamiseen tai sen rajoittamiseen kokonaisvaltaisemmin erilaisia tekniikoita ja tarkoituksia huomioon ottaen. Tässä mielessä fair use -säännöstä voisi DSM-direktiivin 4 artiklan sijaan hedelmällisemmin samaistaa tietoyhteiskuntadirektiivin kappaleen valmistamista koskevaan oikeuteen ja siihen liittyviin poikkeussäännöksiin.

Edellä hahmotellun retoriikan mukaisesti EU:n tekstin- ja tiedonlouhintaa koskevassa tekijänoikeussäätelyssä siirryttiin löyhistä standardeista kohti tarkkarajaisempia sääntöjä. Siinä missä vanha säätely jätti hyvin pienen ja tulkinnanvaraisen mahdollisuuden kaupallisessa tarkoituksessa ilman tekijänoikeuden haltijan lupaa toteutettavalle tekstin- ja tiedonlouhinnalle, DSM-direktiivin 4 artikla antaa eksplisiittisen vapauden louhia tekijänoikeudella suojattua materiaalia niin kauan kuin tekijänoikeuden haltija ei ole nimenomaisesti pidättänyt oikeuksiaan suojatun materiaalin tällaiseen käyttöön. Tätä säätelyllistä siirtymäprosessia voisikin kuvailla siten, että on liikuttu kaupallisessa

¹⁹⁹ Hughes 2017, s. 237.

tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan sallittavuuden häilyvästä mahdollisuudesta tilaan, jossa sallittavuus perustuu joko-tai -tyyppisesti yksinomaan louhittavan aineiston tekijänoikeuden haltijan tahtoon.

Yhdysvalloissa taas tekijänoikeusjärjestelmän ja tekstin- ja tiedonlouhinnan välinen suhde ei ole muuttunut sääntelyllisten muutosten seurauksena. Kohtuullisen käytön doktriini nimittäin on ja pysyy tekijänoikeuksien haltijoiden yksinoikeuksia tietyssä määrin rajoittavana mekanismina. Kuten todettua, fair use -säännöksellä on standardityyppinen luonne, jota voi pitää jossain määrin tietoyhteiskuntadirektiiviä vastaavana. Näin ollen tekstin- ja tiedonlouhintaa ei tekijänoikeudellisesti säännellä erillisenä kokonaisuutena, vaan osana erilaisten kappaleen valmistamisen taustalla olevien tarkoituksien joukkoa. Tällöin ilman tekijänoikeuden haltijan lupaa toteutetun tekijänoikeudella suojatun teoksen louhinnan laillisuusarviointi jää oikeuskäytännön täsmennettäväksi. Google Books -tapaus toimii tällaisena täsmentäjänä: tapauksessa annettujen ratkaisuiden perusteella tekstin- ja tiedonlouhintaa voidaan pitää hyvin transformatiivisena tekijänoikeudella suojatun teoksen käyttönä, minkä seurauksena louhinnan kaupallisella tarkoituksella ei ole arvioinnissa kovinkaan suurta painoarvoa. Myös muiden kohtuullisen käytön kriteereiden voitaneen katsoa melko säännönmukaisesti täyttyvän teoksen käytön transformatiivisuuden seurauksena.

Google Books Library Projectin kaltaisia hankkeita voisi kuvitella olevan melko hankalaa toteuttaa DSM-direktiivin implementoinnin jälkeisessä EU:ssa. Oletus perustuu lähtökohtaan, jonka mukaan Googlen todettaisiin toimivan kaupallisessa tarkoituksessa, jolloin sovellettavaksi tulisi nimenomaan mainitun direktiivin 4 artikla. Mikäli tekijänoikeuksien haltijat pidättäisivät 4 artiklan mukaisesti oikeutensa teoksensa käyttöön, Googlen tai muun vastaavassa asemassa olevan kaupallisen toimijan tulisi hankkia lisenssi kaikkien louhittavaksi tarkoitettujen tekijänoikeudella suojattujen teosten tällaiseen käyttöön. Ottaen huomioon, että esimerkiksi merkittävä osa plagiaatintunnistusjärjestelmistä, internetin hakukoneista ja koneoppimisen tutkimuksista perustuvat kaupalliseen tarkoitukseen²⁰⁰ on hieman kummallista, että EU on päätenyt sääntelyllään ratkaisuun, joka voi potentiaalisesti estää suuren määrän kaupallisessa tarkoituksessa toteutettavista tekstin- ja tiedonlouhintaprojekteista.

²⁰⁰ Sag 2019, s. 323.

5. Johtopäätöksiä

Tässä tutkielmassa on tarkasteltu tekstin- ja tiedonlouhinnan ja tekijänoikeuden välistä suhdetta yhtäältä EU:n sisäisen lainsäädäntökehityksen kannalta ja toisaalta verraten tätä kehityssuuntaa Yhdysvalloissa ajan saatossa muodostuneeseen relevanttiin sääntelyyn ja oikeuskäytäntöön. Tutkielmasta pyrittiin rakentamaan koherentti ja jäsennelty kokonaisuus, jossa teknologinen ja juridinen substanssi kulkevat käsi kädessä ymmärrettävällä tavalla. Tutkielman aiemmissa luvuissa on käyty jo jonkin verran johtopäätöksiksikin kutsuttavissa olevaa pohdintaa sekä siitä, miten EU-lainsäädäntö on DSM-direktiivin myötä muuttunut suhteessa tekstin- ja tiedonlouhintaan että siitä, miten eurooppalainen ja yhdysvaltalainen näkökulma kohtaavat toisensa tässä aihekokonaisuudessa. Käsillä olevassa pääluvussa tullaan syventämään tätä pohdintaa ja esittämään sen pohjalta hieman aiempaa pidemmälle vietyjä tulkintoja. Tässä yhteydessä tuodaan esiin myös joitakin huomioita tekijänoikeudellisesta yleisen ja yksityisen edun välisestä tasapainottelusta nimenomaan tekstin- ja tiedonlouhinnan sääntelyn kontekstissa.

EU-oikeudellisessa viitekehyksessä tekstin- ja tiedonlouhintaa koskevan tekijänoikeussääntelyn voi katsoa sekä muuttuneen että pysyneen tietyssä mielessä ennallaan. Tietysti on niin, että DSM-direktiivi on tuonut mukanaan jäsenvaltioille pakollisesti implementoitaviksi asetetut kappaleen valmistamista koskevaan oikeuteen kajoavat poikkeukset tai rajoitukset. Tämä on vieläpä toteutettu niin ei-kaupallisen kuin kaupallisenkin tarkoituksensa omaavan tekstin- ja tiedonlouhinnan osalta. Kuten tutkielmassa on aiemmin todettu, ei-kaupallisessa tarkoituksessa toteutettavaa tekstin- ja tiedonlouhintaa koskeva DSM-direktiivin 3 artikla on kuitenkin sisällöltään teosten käyttäjille myönteisempi kuin 4 artikla. Voitaneenkin sanoa, että tekstin- ja tiedonlouhinnan EU-tasoinen tekijänoikeussääntely on kokenut konkreettisimman muutoksensa nimenomaan ei-kaupallisessa tarkoituksessa toteutettavan louhinnan saralla.

DSM-direktiivin 4 artiklan 3 kohta on pääasiallinen syy siihen, miksi tekstin- ja tiedonlouhinnan tekijänoikeussääntelyn voi katsoa pysyneen konkreettisesti myös osin ennallaan lainsäädäntömuutosten keskellä. 4 artiklan 3 kohta antaa tutkielmassa aiemmin kuvatulla tavalla oikeudenhaltijoille mahdollisuuden pidättää oikeutensa 4 artiklan 1 kohdassa tarkoitettuun teosten ja muun suojatun aineiston käyttöön. Tämä opt out -mekanismi on sisällytetty ainoastaan DSM-direktiivin 4 artiklaan, joten se koskee vain kaupallisessa tarkoituksessa toteutettavaa tekstin- ja tiedonlouhintaa. Käyttämällä opt out -

mekanismeja aktiivisesti oikeudenhaltijat voivat tehokkaasti rajoittaa 4 artiklan mukaisen poikkeuksen tai rajoituksen käytännön soveltamisen.

On nähdäkseni jo itsessään kyseenalaista luoda sääntelyä, jonka käytännön soveltaminen jää riippuvaiseksi yhden eturyhmän tekemistä valinnoista. Tämä korostuu etenkin silloin, kun tällä eturyhmällä voitaneen olettaa olevan taloudellinen kannustin tehdä nimenomaan sääntelyn käytännön soveltamisen jähmettävä valinta. Louhittavaksi suunniteltujen sisältöjen tekijänoikeuksien haltijoilla mainitun tapaisena kannustimena voisi nähdäkseni toimia se, että opt out -mekanismeja hyödyntämällä oikeudenhaltijoiden olisi mahdollista monetisoida sekä pääsy sisältöön että tämän sisällön louhiminen.

Kolikon kääntöpuolena ovat luonnollisesti tekijänoikeuden syvimpään ytimeen liittyvät lähtökohdat teoksen tekijälle kuuluvista yksinoikeuksista ja niiden ylläpitämisen tärkeydestä luovaan työhön kannustamisessa. Jo valistuksen aikakaudella esitettiin ajatuksia aineettoman omaisuuden olemassaolosta siinä merkityksessä, että teoksen tekijälle tulisi antaa mahdollisuus nauttia työnsä hedelmistä.²⁰¹ Ajatusten taustalla eivät kuitenkaan vaikuttaneet täysin individualistiset pyrkimykset, vaan myös ympäröivää yhteiskuntaa huomioiva tarkoitusperä: teosten tekijät halusivat levittää filosofisia ideoitaan yleisön luettaviksi.²⁰² Tätä kautta voidaan hahmottaa rakenne, jossa tekijänoikeuden olemassaolo ja sitä kautta tekijöiden luovan työn suojaaminen sekä taloudellinen kompensoiminen saavat oikeutuksensa yleisön tarpeesta intellektuellille keskustelulle ja uusille ideoille. Tuo tarve puolestaan vaatii tiettyjen rajoitusten asettamista tekijänoikeuden tuottamalle yksinoikeudelle. Näin ollen yleistä etua voidaan pitää sekä tekijänoikeussuojaa oikeuttavana että sitä rajoittavana perusteena.²⁰³

Edellä sanotusta huolimatta yleisen ja yksityisen edun välisen tasapainon saavuttaminen ei ole osoittautunut helpoksi tehtäväksi. Haaste on relevantti tekijänoikeuden ohella myös yleisemmin immateriaalioikeuden saralla.²⁰⁴ Tekijänoikeusjärjestelmät ympäri maailman on kehitetty ja suunniteltu pyrkien tasapainoon teosten tekijöiden taloudellisten etunäkökohtien ja yleisölle teosten käytöstä syntyvän hyvinvoinnin välillä.²⁰⁵ Lieneekin perusteltua todeta,

²⁰¹ Geiger 2006, s. 367.

²⁰² Geiger 2006, s. 367.

²⁰³ Geiger 2006, s. 367.

²⁰⁴ Oesch 2017, s. 6-7.

²⁰⁵ Hua 2014, s. 39.

että motiivit tekijän ja yleisön etujen tasapainottamiselle eivät ole merkittävästi muuttuneet ajan saatossa.

Pihlajarinne on katsonut, että yleisen ja yksityisen edun välinen tasapainottelu on entisestään hankaloitunut yhteiskunnan digitalisoitumisen seurauksena, minkä lisäksi on argumentoitu, että digitalisoitumiskehitys on asettanut suurimman haasteen kaikkien aineettomien oikeuksien joukosta juuri tekijänoikeudelle.²⁰⁶ Nähdäkseni väitteet ovat perusteltuja jo yksinään piratismiin tekijänoikeuden suomille yksinoikeuksille luomien ongelmien nojalla. Näen piratismiin kuitenkin toimintana, jonka tekijänoikeuden perustavia arvoja vastustavasta luonteesta ja rangaistavuudesta ollaan yleisesti katsottuna melko yksimielisiä. Niinpä yhteiskunnan digitalisoitumisen tekijänoikeudelliselle yleisen ja yksityisen edun väliselle tasapainottelulle asettamat haasteet on sijoitettava laajempaan kontekstiin. Tällöin huomataan, että tasapainottelun lähtökohdat eivät itsessään ole muuttuneet, mutta niiden konkreettiset ilmenemismuodot ovat siirtyneet digitaaliseen toimintaympäristöön.

Edellä esitetyn näkemyksen voi muotoilla myös niin, että tasapainottelun taustalla olevat eri etunäkökohdat ovat voimistuneet digitaalisella aikakaudella.²⁰⁷ Näin ollen kompromissien löytäminen ja sitä kautta tasapainoisten lainsäädännöllisten ratkaisuiden muotoileminen on nykyaikana entistä hankalampaa. Onkin esitetty, että yleiskatsaus kansainvälisistä immateriaalioikeussopimuksista ja vaihtelevien valtioiden kansallisissa tekijänoikeuslaeissa vallitsevista painotuksista osoittaa tekijänoikeuden yksityisenä omaisuutena hahmottavan näkemyksen olevan monessa suhteessa niskan päällä.²⁰⁸ On sanomattakin selvää, että tällainen näkemys on nimenomaan oikeudenhaltijoiden etua korostava ja edustaa nähdäkseni eräänlaista tekijänoikeudellista protektionismia.

Tietoyhteiskuntadirektiivin johdanto-osan 31 perustelukappaleessa todetaan seuraavasti: ”Eri oikeudenhaltijaryhmien välisten samoin kuin eri oikeudenhaltijaryhmien ja suojatun aineiston käyttäjien välistenoikeuksien ja etujen oikeudenmukainen tasapaino on turvattava. Jäsenvaltioissa säädettyjä nykyisiä poikkeuksia ja rajoituksia näihin oikeuksiin on arvioitava uudelleen sähköisen ympäristön kannalta. Tiettyjä yksinoikeustoimia koskevien poikkeusten ja rajoitusten nykyisillä eroilla on välitön kielteinen vaikutus sisämarkkinoiden toimintaan tekijänoikeuksien ja lähioikeuksien alalla. Nämä erot saattavat

²⁰⁶ Pihlajarinne 2004, luku V: Immateriaalioikeudet.

²⁰⁷ Hua 2014, s. 40.

²⁰⁸ Hua 2014, s. 40.

tulla näkyvämmiksi rajojen ylitse tapahtuvan teosten hyödyntämisen ja rajat ylittävien toimien edelleen lisääntyessä. Sisämarkkinoiden asianmukaisen toiminnan varmistamiseksi nämä poikkeukset ja rajoitukset olisi määriteltävä yhdenmukaisemmin. Yhdenmukaistamisen asteen olisi perustuttava vaikutukseen, joka poikkeuksilla ja rajoituksilla on sisämarkkinoiden moitteettomaan toimintaan.”

Tietoyhteiskuntadirektiivin johdanto-osan 31 perustelukappaleesta käy etutasapainotteluun liittyvien tavoitteiden ohella selkeästi ilmi direktiivin EU-tasoiseen tekijänoikeussäätelyn harmonisointiin liittyvä päämäärä. Kuten edellä on todettu, harmonisointi on jäänyt ainakin tietoyhteiskuntadirektiivin nojalla hieman vajavaiseksi.

Tietokantadirektiiviin ei ole sisällytetty tietoyhteiskuntadirektiivin johdanto osan 31 perustelukappaletta vastaavaa kappaletta. Tietokantadirektiivin johdanto-osan 45 perustelukappale liittyy kuitenkin tiiviisti eri intressitahojen välillä tehtävään tasapainotteluun. Kyseisessä kappaleessa lausutaan seuraavaa: ”Oikeus estä luvaton kopiointi ja/tai uudelleenkäyttö ei tarkoita millään tavalla tekijänoikeudellisen suojan ulottamista pelkkiin tosiasioihin tai tietoihin.” Tätä tavoitetta voinee nykypäivänä tarkastella hieman kriittisestikin, sillä jo yli kymmenen vuotta sitten on esitetty huoli kehityssuunnasta, jossa tietoa ”yksityistetään” antamalla sille suojaa ainoastaan omana itsenään, kuten esimerkiksi tietokantoja koskevan sui generis -oikeuden osalta on toimittu.

DSM-direktiivin johdanto-osan 6 perustelukappaleessa todetaan puolestaan seuraavasti: ”Tässä direktiivissä säädetyillä poikkeuksilla ja rajoituksilla pyritään saavuttamaan oikeudenmukainen tasapaino yhtäältä tekijöiden ja muiden oikeudenhaltijoiden ja toisaalta käyttäjien etujen ja oikeuksien välillä. Niitä voidaan soveltaa ainoastaan tietyissä erityistapauksissa, jotka eivät ole ristiriidassa teosten tai muun suojatun aineiston tavanomaisen hyödyntämisen kanssa eivätkä kohtuuttomasti haittaa oikeudenhaltijan oikeutettuja etuja.” Perustelukappaleen jälkimmäinen virke vastaa tietoyhteiskuntadirektiivin 5 artiklan 5 kohtaa, joka on puolestaan jo edellä todetulla tavalla omaksuttu niin kutsutusta kolmivaihetestistä.

Esitellyt eri direktiivien perustelukappaleet tukevat näkemystä siitä, että tekstin- ja tiedonlouhintaa koskeva käytännön tekijänoikeuskehikko saattaa tulla DSM-direktiivinkin jälkeen muuttumaan vain tietyiltä osin. Tietoyhteiskuntadirektiivi ja tietokantadirektiivi sisältävät tavoitteita yleisen ja yksityisen edun välisen tasapainon saavuttamisesta, ja

tavoitteita on pyritty konkretisoimaan direktiiveihin laadittujen poikkeusten tai rajoitusten avulla. Kuten tutkielmassa on havainnollistettu, tässä ei ole tekstin- ja tiedonlouhinnan osalta onnistuttu, oli kyse sitten ei-kaupallisesta tai kaupallisesta tarkoituksessa toteutettavasta louhinnasta. Syy epäonnistumiseen lienee ainakin ei-kaupallisesta tarkoituksessa toteutettavan louhinnan osalta pääosin siinä, että poikkeuksia ja rajoituksia ei ole laadittu jäsenvaltioille pakollisesti implementoitaviksi. DSM-direktiivin 3 artikla kuitenkin poistaa tältä osin tekstin- ja tiedonlouhinnan tekijänoikeudelliseen oikeutukseen liittyvää epäselvyyttä ja muuttaa siten nähdäkseni myös yleisen ja yksityisen edun balanssia enemmän tavoitteitaan vastaavaksi.

Kaupallisen tekstin- ja tiedonlouhinnan osalta tietoyhteiskuntadirektiivin ja tietokantadirektiivin sisältämät poikkeukset ja rajoitukset on nähty tutkielmassa aiemmin kuvatulla tavalla melko hyödyttömiksi. Näin ollen voidaan tulkita, että lainsäätäjä ei ole tuolloin nähnyt erityistä tarvetta tasapainottaa yleistä ja yksityistä etua kaupallisesta tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan osalta erillisen poikkeuksen tai rajoituksen muodossa. Tilanne on toki tältä osin melko lailla vastaava myös ei-kaupallisen louhinnan osalta. Louhinnan tarkoituksien yleisen ja yksityisen edun väliseen tasapainotteluun tuomat erot näkyvät kuitenkin DSM-direktiivissä, kun direktiivin 3 artiklassa ei anneta 4 artiklan tapaista mahdollisuutta oikeudenhaltijoille pidättää oikeuttaan teoksen käyttöön tekstin- ja tiedonlouhinnassa. Direktiivin johdanto-osan 17 perustelukappaleessa todetaan vieläpä, että 3 artiklan sisältämän poikkeuksen luonne ja laajuus eli se, että poikkeus rajoittuu tieteellistä tutkimusta harjoittaviin yhteisöihin tarkoittaa sitä, että poikkeuksesta oikeudenhaltijoille syntyvä mahdollinen haitta on erittäin vähäinen.

Nähdäkseni edellä kuvattu problematiikka tarkoittaa yksinkertaistettuna sitä, että yleisen ja yksityisen edun tasapainottamisessa rajat on vedetty eri kohtiin DSM-direktiivin 3 ja 4 artikloissa. Tämä saattaa johtaa tilanteeseen, jossa ei-kaupallisesta tarkoituksessa toteutettava tekstin- ja tiedonlouhinta on saanut kaipaamansa lainsäädäntömuutoksen, joka toimii myös käytännössä, kun taas sen kaupallinen vastine jää tätä vaille. Juuri tämä kontrasti perusteli tilanteen vertaamista Yhdysvaltojen tekstin- ja tiedonlouhintaa koskevaan sääntelyyn ja oikeuskäytäntöön. Tämän tarkastelun pohjalta Yhdysvaltojen mallia olisi voitu harkita sovellettavaksi myös Euroopassa.

Yhdysvalloissa tekijänoikeuden rajoitukset ja poikkeukset eivät perustu niinkään yksittäisiin säännöksiin, vaan ennemmin kohtuullisen käytön periaatteen tapauskohtaiseen soveltamiseen. Yleisellä tasolla tällaisen lähestymistavan voi arvioida tuottavan sekä positiivisia että myös joitakin negatiivisia seurauksia tekijänoikeusjärjestelmän tehokkaalle toimimiselle. Tapauskohtaisesti joustava poikkeus- tai rajoitussäännös on nähdäkseni huomattavasti teknologianeutraalimpi ja aikaa kestävämpi ratkaisu kuin erikseen lainsäädäntöuudistusten kautta säädettävät yksittäiset tietyn teknologian hyödyntämiseen perustuvat säännökset. Toisaalta tapauskohtainen joustavuus tuonee mukanaan myös tiettyä ennakoimattomuutta ja epävarmuutta verrattuna tilanteeseen, jossa poikkeukset ja rajoitukset olisi kirjattu tarkkarajaisesti osaksi lainsäädäntöä. Tästä syystä ei ole nähdäkseni mahdollista todeta yhtä järjestelmää yksiselitteisesti paremmaksi kuin toista.

Tämän tutkielman tarkoitus ei ole kuitenkaan tehdä yleisen tason johtopäätöksiä siitä, onko Yhdysvaltojen tekijänoikeusjärjestelmä toimivampi vastaamaan kaikkiin digitalisaation asettamiin haasteisiin paremmin kuin EU:n hieman pirstaloituneempi sääntelykehikko. Tutkielmassa tehdyn tarkastelun perusteella on sen sijaan mahdollista vetää joitakin johtopäätöksiä siitä, kumpi järjestelmästä tukee vahvemmin kaupallisessa tarkoituksessa toteutettavaa tekstin- ja tiedonlouhintaa. Tässä vertailussa Yhdysvaltojen voitaneen katsoa päätyneen digitalisaatiokehityksen kannalta myönteisempään ratkaisuun, sillä fair use -säännökseen perustuva järjestelmä on muun muassa Google Books -tapauksen kautta vahvistanut, että kaupallisessa tarkoituksessa toteutettava tekstin- ja tiedonlouhinta voi tulla yhtä lailla luetuksi tekijänoikeuden alaisen teoksen kohtuullisen käytön piiriin kuin sen ei-kaupallisessa tarkoituksessa toteutettava vastine. EU:ssa louhinnan tarkoituksellisuuden kaupallisuus sen sijaan on todettuna tavalla merkinnyt DSM-direktiivin 4 artiklan 3 kohdan mukaisen ylimääräisen ”esteen” luomista.

Digitalisaatiokehityksen edistäminen, koneoppimisen mahdollistaminen tai yleinen etu ylipäänsä eivät kuitenkaan ole ainoita tekijöitä, jotka tulisi ottaa huomioon pohdittaessa kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan solauttamista osaksi tekijänoikeuden rajoitusten ja poikkeusten listaa. Punnitsemisessa vastapainona tulisi aina huomioda myös yksityinen etu ja tekijöiden kannustaminen luovaan työhön. Onkin luonnollisesti niin, että tekijänoikeuden merkitys menetettäisiin tyystin, jos sen luomista yksinoikeuksista voitaisiin poiketa ainoastaan vetoamalla erinäisiin yhteiskunnan kehittymistä ja yleistä etua korostaviin näkökohtiin.

Kaupallisessa tarkoituksessa toteutettava tekstin- ja tiedonlouhinta ei nähdäkseni kuitenkaan vaaranna louhittaviksi suunniteltujen teosten tekijöiden asemaa oikeudenhaltijoina. Ei ainakaan esimerkiksi silloin, kun louhinnan tarkoituksena on syöttää harjoitusdataa koneelle ja auttaa sitä näin tunnistamaan syötetystä aineistosta kiinnostavia malleja, joiden perusteella voidaan tehdä uusia havaintoja. Tällöin on kyse käytöstä, joka eroaa alkuperäisen teoksen käytöstä sekä käyttötavaltaan että -tarkoitukseltaan, mikä taas on kohtuullisen käytön doktriinissa transformatiivisen teoksen käytön ja käyttötarkoituksen ydin. Tällainen transformatiivisuus tekstin- ja tiedonlouhintaprosessin luonteessa johtaa nähdäkseni myös tulkintaan siitä, ettei prosessi uhkaa korvata alkuperäistä teosta saatikka kilpaile sen kanssa, vaan ainoastaan tuottaa siitä lisää tietoa.

Nähdäkseni kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan EU-tasoinen sääntely olisi voinut uudistuessaan omaksua asioita Yhdysvaltojen fair use -traditioon liittyvästä arvioinnista. Tällöin yleisen ja yksityisen edun välinen punninta olisi voinut tulla paremmin huomioiduksi, eikä kaupallisessa tarkoituksessa toteutettavan tekstin- ja tiedonlouhinnan suorittamiselle tekijänoikeuden rajoituksen tai poikkeuksen nojalla olisi tarvinnut asettaa opt out -mekanismin kaltaista kategorista estettä. Ratkaisulla lienee joitakin oikeusvarmuutta ehostavia vaikutuksia, mutta sillä on myös potentiaalia tarpeettomasti jarruttaa eurooppalaista tekoälykehitystä ylikorostamalla oikeudenhaltijoiden etua yleisen edun kustannuksella.